Balanced Allocations:
Caching and Packing, Twinning and Thinning*

Dimitrios Los[1], Thomas Sauerwald[†1], and John Sylvester[‡2]

[1]Department of Computer Science & Technology, University of Cambridge, UK,
`firstname.lastname@cl.cam.ac.uk`
[2]School of Computing Science, University of Glasgow, UK,
`john.sylvester@glasgow.ac.uk`

October 20, 2021

Abstract

We consider the sequential allocation of m balls (jobs) into n bins (servers) by allowing each ball to choose from some bins sampled uniformly at random. The goal is to maintain a small gap between the maximum load and the average load.

In this paper, we present a general framework that allows us to analyze various allocation processes that slightly prefer allocating into underloaded, as opposed to overloaded bins. Our analysis covers several natural instances of processes, including:

- The Caching process (a.k.a. memory protocol) as studied by Mitzenmacher, Prabhakar and Shah (2002): At each round we only take one bin sample, but we also have access to a cache in which the most recently used bin is stored. We place the ball into the least loaded of the two.
- The Packing process: At each round we only take one bin sample. If the load is below some threshold (e.g., the average load), then we place as many balls until the threshold is reached; otherwise, we place only one ball.
- The Twinning process: At each round, we only take one bin sample. If the load is below some threshold, then we place two balls; otherwise, we place only one ball.
- The Thinning process as recently studied by Feldheim and Gurel-Gurevich (2021): At each round, we first take one bin sample. If its load is below some threshold, we place one ball; otherwise, we place one ball into a *second* bin sample.

As we demonstrate, our general framework implies for all these processes a gap of $\mathcal{O}(\log n)$ between the maximum load and average load, even when an arbitrary number of balls $m \geqslant n$ are allocated (heavily loaded case). Our analysis is inspired by a previous work of Peres, Talwar and Wieder (2010) for the $(1+\beta)$-process, however here we rely on the interplay between different potential functions to prove stabilization.

Keywords— Balls in bins, balanced allocations, potential functions, heavily loaded, gap bounds, maximum load, thinning, two-choices, weighted balls.
AMS MSC 2010— 68W20, 68W27, 68W40, 60C05

*Full version of a paper appearing in SODA 2022

†T.S. was supported by the ERC Starting Grant 679660 (DYNAMIC MARCH).
‡J.S. was supported by EPSRC project EP/T004878/1: Multilayer Algorithmics to Leverage Graph Structure

Contents

1 Introduction

We consider the sequential allocation of m balls (jobs or data items) to n bins (servers or memory cells), by allowing each ball to choose from a set of randomly chosen bins. The goal is to allocate balls efficiently, while also keeping the load distribution balanced. The balls-into-bins framework has found numerous applications in hashing, load balancing, routing, but has also shown to be nicely related to more theoretical topics such as randomized rounding or pseudorandomness (we refer to the surveys [30] and [22] for more details).

A classical algorithm is the d-CHOICE process introduced by Azar et al. [1] and Karp et al. [17], where for each ball to be allocated, we sample $d \geqslant 1$ bins uniformly and then place the ball in the least loaded of the d sampled bins. It is well-known that for the ONE-CHOICE process ($d = 1$), the gap between the maximum and average load is $\Theta\big(\sqrt{m/n \cdot \log n}\big)$. In particular, this gap grows significantly as $m/n \to \infty$, which is called the *heavily loaded case*. For $d = 2$, [1] proved that the gap is only $\log_2 \log n + \mathcal{O}(1)$ for $m = n$. This result was generalized by Berenbrink et al. [2] who proved that the same guarantee also holds for $m \geqslant n$, in other words, even as $m/n \to \infty$, the difference between the maximum and average load remains a slowly growing function in n that is independent of m. This dramatic improvement of TWO-CHOICE over ONE-CHOICE has been widely known as the "power of two choices".

It is natural to investigate allocation processes that are less powerful than TWO-CHOICE in their ability to sample two bins, sample uniformly or distinguish between the load of the sampled bins. Such processes make fewer assumptions than TWO-CHOICE and can thus be regarded as more sampling-efficient and robust.

One example of such an allocation process is the $(1 + \beta)$-process introduced by Peres et al. [24], where each ball is allocated using ONE-CHOICE with probability $1 - \beta$ and otherwise is allocated using TWO-CHOICE. The authors proved that for any $\beta \in (0, 1]$, the gap is only $\mathcal{O}(\frac{\log n}{\beta})$. Hence, only a "small" fraction of TWO-CHOICE rounds are enough to inherit the property of TWO-CHOICE that the gap is independent of m. A similar class of adaptive sampling schemes (where, depending on the loads of the samples so far, the ball may or may not sample another bin) was analyzed by Czumaj and Stemann [6], but their results hold only for $m = n$.

Another important example is 2-THINNING (which we later simply refer to as THINNING), which was studied in [10, 16]. In this process, each ball first samples a bin uniformly and if its load is below some threshold, the ball is placed into that sample. Otherwise, the ball is placed into a *second* bin sample, without inspecting its load. In [10], the authors proved that for $m = n$, there is a fixed threshold which achieves a gap of $\mathcal{O}(\sqrt{\frac{\log n}{\log \log n}})$. This is a significant improvement over ONE-CHOICE, but also the total number of samples is $(1+o(1))\cdot m$, which is an improvement over TWO-CHOICE. Similar threshold processes have been studied in queuing [9], [20, Section 5] and discrepancy theory [8]. For values of m sufficiently larger than n, [11] and [18] prove some lower and upper bounds for a more general class of *adaptive* thinning protocols (here, adaptive means that the choice of the threshold *may* depend on the load configuration). Related to this line of research, [18] also analyze a so-called QUANTILE-process, which is a version of THINNING where the ball is placed into a second sample only if the first bin has a load which is at least the median load.

Finally, we mention the CACHING process (also known as memory-protocol) analyzed by Mitzenmacher et al. [21], which is essentially a version of the TWO-CHOICE process with cache. At each round, we take a uniform sample but we also have access to a cache. Then the ball is placed in the least loaded of the sampled and the bin in the cache, and after that the cache is updated if needed. It was shown in [21] that for $m = n$, the process achieves a better gap than TWO-CHOICE.

From a more technical perspective, apart from analyzing a large class of natural allocation processes, an important question is to understand how sensitive the gap is to changes in the allocation distribution. To this end, [24] formulated general conditions on the distribution vector, which, when satisfied in all rounds, imply a small gap bound. These were then later applied not only to the $(1+\beta)$-process, but also to analyze a "graphical" allocation model where a pair of bins is sampled by picking an edge of a graph uniformly at random. Other works which study perturbations on the allocation distribution are: allocations on hypergraphs [15], balls-into-bins with correlated choices [29]; or balls-into-bins with hash functions [4].

Our Results. We introduce a general framework that allows us to deduce a small gap independent of m for many of the processes above, but also opens the avenue for the study of novel allocation processes such as TWINNING or PACKING. To ensure that an allocation process produces a balanced load distribution, we could either bias the allocation towards underloaded bins by skewing the probability by which a bin is chosen for an allocation (probability bias), or alternatively, we could add more balls to a bin if it is underloaded (weight bias).

Note that a small bias in the probability distribution can be achieved in various ways, for example, by taking a *second* bin sample: (*i*) in each round (TWO-CHOICE), or (*ii*) in each round with some probability β ($(1+\beta)$-process), or (*iii*) in each round dependent on the load of the first sample (THINNING). Regarding the weight bias, as pointed out in [21], allocating consecutive balls to the same bin is very natural in applications like sticky routing [14] or load balancing, where tasks (=balls) may often arrive in bursts. We note that the filling bias can be also seen as a "popularity-bias"; once a lightly loaded bin has been found, it will be a preferred location for the next balls until it becomes overloaded.

The first type of allocation processes matching the above definition are called **Filling Processes**, where it suffices to sample one uniform bin for each allocation, but then we are able to "fill" the bin with balls until it becomes overloaded:

> **Condition $\mathcal{P}_1$:** At each round, the probability of allocating to any fixed bin is majorized by the uniform distribution (i.e., ONE-CHOICE).
> **Condition $\mathcal{W}_1$:** At each round, if an underloaded bin is chosen, allocate just enough balls to it so that the bin becomes overloaded. Otherwise, place a single ball.

As our first main result, we prove that if $\mathcal{P}_1$ *and* a (more technical, but relaxed) version of $\mathcal{W}_1$ both hold, then a gap bound of $\mathcal{O}(\log n)$ follows. While it is easy to see that PACKING meets the two conditions, some care is needed to apply the framework to CACHING (a.k.a. memory-protocol) due to its use of the cache, which creates correlations between the allocation of any two consecutive balls. Hence PACKING and CACHING are at least as sampling-efficient as ONE-CHOICE while achieving a gap independent of m, which matches the gap bound of $(1+\beta)$-process which requires strictly more samples per ball.

The second type of allocation processes matching the above description are called **Non-Filling Processes**. In these processes, we can only allocate a constant number of balls to a bin, regardless of its load. Neglecting some technical details (see Section 4.2), the two conditions are as follows:

> **Condition $\mathcal{P}_3$:** At each round, the probability of allocating to any fixed underloaded bin is larger than $\frac{1}{n}$, while the probability for any fixed overloaded bin is smaller than $\frac{1}{n}$.
> **Condition $\mathcal{W}_3$:** At each round, if an underloaded bin is chosen for allocation, place more (but a constant number) of balls than if an overloaded bin is chosen.

As our second main result, we prove a gap bound of $\mathcal{O}(\log n)$ whenever a process satisfies *at least one* of these two conditions. It turns out that $\mathcal{P}_3$ is satisfied by MEAN-THINNING (i.e., the 2-Thinning process with threshold being the average load), while $\mathcal{W}_3$ is satisfied by TWINNING; thus for both of these processes a gap bound of $\mathcal{O}(\log n)$ follows immediately. The result for TWINNING gives an example of a process with gap $\mathcal{O}(\log n)$, which (*i*) samples one bin in each round and (*ii*) allocates at most a constant number of balls in any round. The result for MEAN-THINNING extends via an inductive argument to any THINNING process with a threshold between 0 and $\mathcal{O}(\log n)$ above the mean load. Finally, using the idea of majorizing allocation probabilities [24], our framework also applies to $(1+\beta)$-processes for constant $\beta \in (0,1]$.

To the best of our knowledge, most of the related work on balls-into-bins focuses on a small number (usually one or two) allocation processes, and analyzes their gap (with [24, 28] being exceptions). Here we develop a framework that captures a variety of existing and new processes. This generalization comes at a price, e.g., the $\mathcal{O}(\log n)$ bound for Caching is probably not tight and a more specialized analysis may be needed for an improved bound. Nonetheless, for many processes our upper bounds are tight up to constant factors (or $\mathcal{O}(\log\log n)$), see Table 1.

Our analysis, especially of Twinning, bears some resemblance to the weighted balls-into-bins setting. There, the weights of the balls are usually drawn from some distribution *before* the bin is sampled (see [3, 24, 27] for some results). The major difference is that in our model weights depend on the chosen bin, whereas in previous works this is not the case, but instead there is a time-invariant probability bias towards lightly loaded bins (e.g., as in $(1+\beta)$-process and Two-Choice).

Process	Lightly Loaded Case $m = \mathcal{O}(n)$		Heavily Loaded Case $m = \omega(n)$	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
$(1+\beta)$, fixed $\beta \in (0,1)$	$\frac{\log n}{\log\log n}$ [24]		$\log n$	$\log n$
Caching	$\log\log n$ [21]		–	$\log n$
Packing	$\frac{\log n}{\log\log n}$		$\frac{\log n}{\log\log n}$	$\log n$
Twinning	$\frac{\log n}{\log\log n}$		$\log n$	
Mean-Thinning	$\frac{\log n}{\log\log n}$		$\log n$	
Thinning $\left(\Theta(\sqrt{\frac{\log n}{\log\log n}})\right)$	$\sqrt{\frac{\log n}{\log\log n}}$ [12]		$\frac{\log n}{\log\log n}$ [18]	$\log n$
Adaptive-Thinning	$\sqrt{\frac{\log n}{\log\log n}}$ [12]		$\frac{\log n}{\log\log n}$ [18]	$\frac{\log n}{\log\log n}$ [11]

Table 1: Overview of the Gap achieved (with probability at least $1-n^{-1}$), by different allocation processes considered in this work. All stated bounds hold asymptotically; upper bounds hold for all values of m, while lower bounds may only hold for certain values of m. Cells shaded in Green are new results, cells shaded Gray are known results we re-prove. The upper bounds for Packing and Twinning when $m = \mathcal{O}(n)$ follow immediately from One-Choice.

Proof Overview. To analyze filling processes, we consider an exponential potential function Φ^t with parameter $\alpha > 0$ (a variant of the one used in [24, 26]). This potential function considers only bins that are overloaded by at least two balls; thus it is blind to the load configuration across underloaded bins. Our first lemma (Lemma 5.1) upper bounds $\mathbf{E}\left[\,\Phi^{t+1} \mid \Phi^t\,\right]$, and essentially proves that this expectation is maximized if the process uses a uniform distribution for picking a bin. We then use this upper bound to deduce that: (i) in many rounds, the potential drops by a factor of at least $1-\Theta(\frac{\alpha}{n})$, and (ii) in other rounds, the potential increases by a factor of at most $1+\Theta(\frac{\alpha^2}{n})$ (see Lemma 5.2). Taking α sufficiently small, we conclude that $\mathbf{E}\left[\,\Phi^t\,\right] = \mathcal{O}(n)$, and by using Markov's inequality the desired gap bound is implied.

The analysis of non-filling processes is considerably more involved. The main technical challenge is that for large m, these processes may undergo long phases in which either most of the bins are overloaded or most of the bins are underloaded, and consequently most balls are allocated using One-Choice (see Fig. 6 for some experiments). Note that this is not the case for other processes such as the $(1+\beta)$-process (for constant $\beta > 0$) let alone Two-Choice.

We overcome this challenge by establishing a sufficient "self-stabilisation" of the quantile of the mean load. This is done by first proving that the quadratic potential drops as long as the absolute potential is $\Omega(n)$ (Lemma 6.2). Hence, the absolute potential has to be $\mathcal{O}(n)$ at some point. Then we prove that once this happens, the quantile of the mean load will be stable, i.e., bounded away from 0 and 1, for sufficiently many rounds (Lemma 6.1). The final ingredient is to prove that in rounds were the mean quantile is stable, the exponential potential drops by a sufficiently large factor in expectation, while in other rounds it increases by a smaller factor. This step is similar to the analysis of filling processes, but more complicated; one reason being that the exponential potential function here includes both overloaded and underloaded bins.

Organization. In Section 2 we describe the aforementioned allocation processes more formally (in addition, an illustration of the four processes mentioned in the abstract can be found in Fig. 1 on page 7). In Section 3 we define some basic mathematical notation needed for our analysis. In Section 4 we present our general framework for filling and non-filling processes. In this part, we also verify that the specific processes mentioned in the abstract fall into (or can be reduced to) the framework. In Section 5, we prove an $\mathcal{O}(\log n)$ bound on the gap for filling processes. In Section 6 we sketch the proof of the same gap bound for non-filling processes. In Section 7 we introduce a crucial concept for the analysis of non-filling processes, which are tools that imply that the mean quantile stabilizes. In Section 8, we derive several inequalities involving different kinds of potential functions. In Section 9, we complete the non-filling analysis. The derivation of the lower bound is given in Section 10. We also present experiments (Fig. 6 on page 35) illustrating the interplay between the different potential functions used in the analysis. In Appendix C, we empirically compare the gaps of the different processes.

2 Balanced Allocation Processes

We consider sequential allocation processes that allocate m balls into n bins. By x^t we denote the load vector at round t (the state after t rounds have been completed). Further, $W^t := |x^t|_1$ denotes the total number of balls allocated. We begin with the classical d-Choice process [1, 17].

> d-Choice Process:
> Iteration: For each $t \geqslant 0$, sample d uniform bins $i_1, \ldots, i_d$ independently. Place a ball in a bin i_{min} satisfying $x^t_{i_{\min}} = \min_{1 \leqslant j \leqslant d} x^t_{i_j}$, breaking ties arbitrarily.

Mixing One-Choice with Two-Choice rounds at a rate β, one obtains the $(1+\beta)$ process [24]:

> $(1+\beta)$ Process:
> Parameter: A probability $\beta \in (0, 1]$.
> Iteration: For each $t \geqslant 0$, with probability β allocate one ball via the Two-Choice process, otherwise allocate one ball via the One-Choice process.

The memory protocol was introduced by Mitzenmacher et al. [21] and works under the assumption that the address b of a single bin can be stored or 'cached'. The process is essentially a Two-Choice process where the second sample is replaced by the bin in the cache.

> Caching Process (a.k.a. Memory Protocol):
> Iteration: For each $t \geqslant 0$, sample a uniform bin i, and update its load (or of cached bin b):
>
> $$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t < x_b^t \qquad \text{(also update cache } b = i\text{)}, \\ x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t = x_b^t, \\ x_b^{t+1} = x_b^t + 1 & \text{if } x_i^t > x_b^t. \end{cases}$$

As mentioned earlier, [11, 18] studied a more powerful version of Thinning, where the threshold function may be adapted to the current load vector, i.e., $f = f(n, x^t)$. However, in this work we focus on non-adaptive threshold functions $f(n)$.

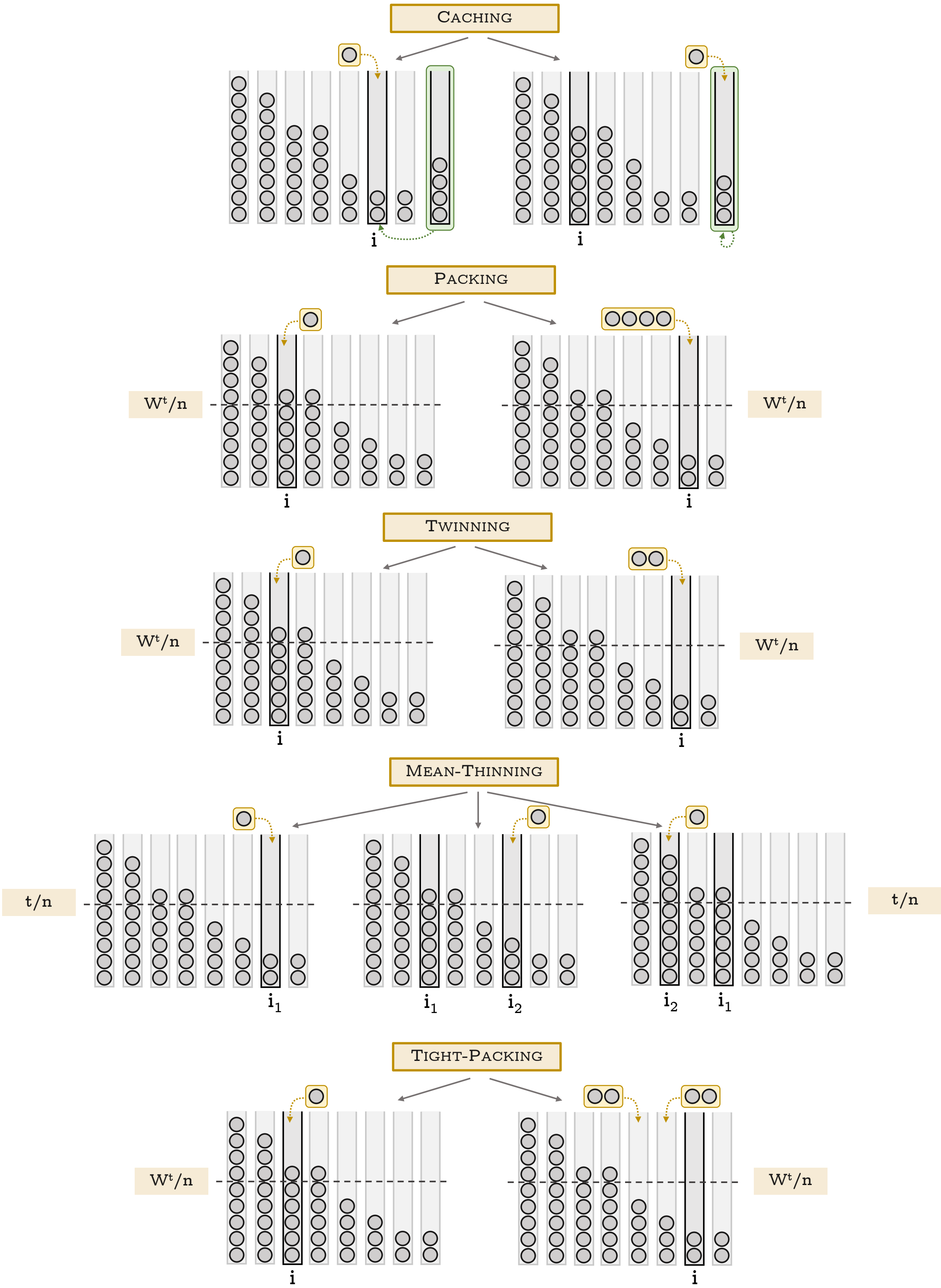

Figure 1: Illustration of the different possibilities of allocating one (or more) balls in each of the four processes: Caching, Packing, Twinning, Mean-Thinning, Tight-Packing.

We further introduce two new processes: (a) the Packing process, which places "greedily" as many balls as possible into an underloaded bin (i.e., a bin with load below average) and (b) the Twinning process, which is similar to Packing, but in a sense it is less extreme as it only places two balls into an underloaded bin. This process is possibly the most similar to

ONE-CHOICE among our processes, however, we still prove that it has a $\mathcal{O}(\log n)$ gap.

<u>PACKING Process:</u>
Iteration: For each $t \geqslant 0$, sample a uniform bin i, and update its load:

$$x_i^{t+1} = \begin{cases} \lceil \frac{W^t}{n} \rceil + 1 & \text{if } x_i^t < \frac{W^t}{n}, \\ x_i^t + 1 & \text{if } x_i^t \geqslant \frac{W^t}{n}. \end{cases}$$

<u>TWINNING Process:</u>
Iteration: For each $t \geqslant 0$, sample a uniform bin i, and update its load:

$$x_i^{t+1} = \begin{cases} x_i^t + 2 & \text{if } x_i^t < \frac{W^t}{n}, \\ x_i^t + 1 & \text{if } x_i^t \geqslant \frac{W^t}{n}. \end{cases}$$

The following process has been studied by several authors [10, 12, 16], where the threshold is usually set in advance. A special case is THINNING(0) which we also call MEAN-THINNING.

<u>THINNING($f(n)$) Process:</u>
Parameter: A threshold function $f(n) \geqslant 0$
Iteration: For $t \geqslant 0$, sample two uniform bins i_1 and i_2 independently, and update:

$$\begin{cases} x_{i_1}^{t+1} = x_{i_1}^t + 1 & \text{if } x_{i_1}^t < \frac{t}{n} + f(n), \\ x_{i_2}^{t+1} = x_{i_2}^t + 1 & \text{if } x_{i_1}^t \geqslant \frac{t}{n} + f(n). \end{cases}$$

The above processes arise naturally when one is trying to achieve a small gap from few bin queries or random samples. In contrast, the following process is rather contrived, as whenever an underloaded bin is chosen, balls can be placed into (possibly different) underloaded bins that have the highest load. However, it is interesting that even this process achieves a small gap, as we will prove later.

<u>TIGHT-PACKING Process:</u>
Iteration: For each $t \geqslant 0$, sample a uniform bin i, and update:

$$\begin{cases} \text{Place } \lceil -x_i^t + \frac{W^t}{n} \rceil + 1 \text{ balls one by one into the bins with highest loads} \\ \text{such that } x_j^{t+1} < \frac{W^t}{n}, \text{ except for one bin } j \text{ with } \frac{W^t}{n} \leqslant x_j^{t+1} < \frac{W^t}{n} + 1, & \text{if } x_i^t < \frac{W^t}{n}, \\ x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t \geqslant \frac{W^t}{n}. \end{cases}$$

An equivalent description of TIGHT-PACKING with decreasingly sorted bin loads $x_1^t \geqslant x_2^t \geqslant \cdots \geqslant x_n^t$ goes as follows. After picking the random bin i, we update the load of the maximally loaded underloaded bin $j \in [n]$ at round t to $x_j^{t+1} = x_j^t + \lceil -x_j^t + \frac{W^t}{n} \rceil$, so its new load satisfies $x_j^{t+1} \in [\frac{W^t}{n}, \frac{W^t}{n} + 1)$. Then the remaining $\lceil -x_i^t + \frac{W^t}{n} \rceil + 1 - \lceil -x_j^t + \frac{W^t}{n} \rceil \geqslant 0$ balls (if there are any), are allocated to bins $j+1, j+2, \ldots, j+\ell$ for some integer $\ell \geqslant 0$, such that all bins $i \in [j+1, j+\ell-1]$ have the same load value in $[\frac{W^t}{n} - 1, \frac{W^t}{n})$.

For illustrations of TIGHT-PACKING and other processes defined in this section see Fig. 1.

3 Notation

As mentioned before, x^t is the load vector of the n bins at round $t = 0, 1, 2, \ldots$ (the state after t rounds have been completed). So in particular, $x^0 := (0, \ldots, 0)$. In many parts of the analysis, we will assume an arbitrary labeling of the n bins so that their loads at round t are ordered decreasingly, i.e.,

$$x_1^t \geqslant x_2^t \geqslant \cdots \geqslant x_n^t.$$

Unlike many of the standard balls-into-bins processes, some of our processes may allocate more than one ball in a single round. To this end we define $W^t := \sum_{i \in [n]} x_i^t$ as the total number of

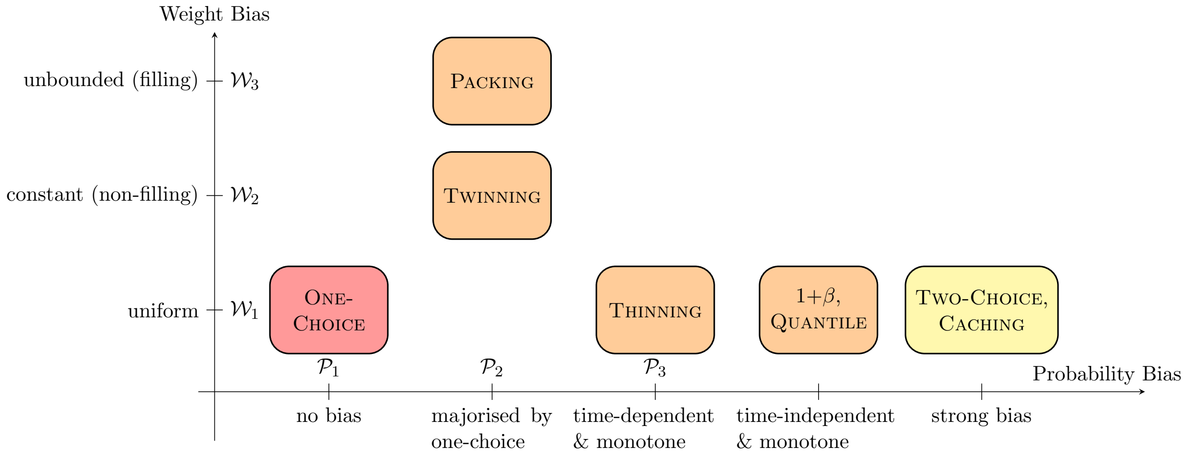

Figure 2: A schematic overview of different allocation processes. For a formal definition of QUANTILE, we refer the reader to [18].

balls allocated that are allocated by round t (for THINNING and CACHING we allocate one ball per round, so $W^t = t$; for PACKING and TWINNING, $W^t \geqslant t$). We will also use $w^t := W^{t+1} - W^t$ to denote the number of balls allocated in round t.

We define the gap as $\text{Gap}(t) := \max_{i \in [n]} x_i^t - \frac{W^t}{n}$, which is the difference between the maximum load and average load[1] at round t. When referring to the gap of a specific process P, we write $\text{Gap}_P(t)$ but may simply write $\text{Gap}(t)$ if the process under consideration is clear from the context. Finally, we define the normalized load of a bin i as:

$$y_i^t := x_i^t - \frac{W^t}{n}.$$

Further, let $B_+^t := \left\{ i \in [n] \colon y_i^t \geqslant 0 \right\}$ be the set of overloaded bins, and $B_-^t := [1, n] \setminus B_+^t$ be the set of underloaded bins. Let $\delta^t := |B_+^t|/n \in [\frac{1}{n}, 1]$ which is the quantile corresponding to the average load. Following [24], let $p_1^t, p_2^t, \ldots, p_n^t$ be the distribution vector of an allocation process, where for every $i \in [n]$, p_i^t is the probability that the process picks bin the i-th most heavily loaded bin round t for the allocation. We denote by $i \in_{p^t} [n]$ a sample of $[n]$ according to this vector p^t. Note that in most processes after taking the sample i, all balls will be allocated to i; however, when we define filling processes in the next section, this requirement will be relaxed. A special case of a distribution vector is the uniform distribution vector of ONE-CHOICE, which is $p_i^t = p_i = \frac{1}{n}$ for all i and t. For two distribution vectors p and q (or analogously, for two sorted load vectors), we say that p majorizes q if for all $1 \leqslant k \leqslant n$, $\sum_{i=1}^k p_i \geqslant \sum_{i=1}^k q_i$. Let $P_+^t := \sum_{i \in B_+^t} p_i^t$ and $P_-^t := \sum_{i \in B_-^t} p_i^t$.

In the following, $\mathfrak{F}^t$ is the filtration corresponding to the first t allocations of the process (so in particular, $\mathfrak{F}^t$ reveals x^t). For random variables Y, Z we say that Y is stochastically smaller than Z (or equivalently, Y is stochastically dominated by Z) if $\mathbf{Pr}\left[Y \geqslant x \right] \leqslant \mathbf{Pr}\left[Z \geqslant x \right]$ for all real x.

Throughout the paper, we often make use of statements and inequalities which hold only for sufficiently large n. For simplicity, we do not state this explicitly.

4 General Framework and Upper Bounds

We present two general frameworks that cover allocation processes with a certain bias towards underloaded bins. The first framework is called "filling"-processes, as such processes will be

[1] It is common in the literature to focus on this difference, rather than the difference between maximum and minimum load; however, our results for non-filling extend to this stronger notion.

allowed to fill underloaded bins to their maximum capacity (which is defined to be the average load plus one). The second framework covers so-called "non-filling"-processes, where the number of balls that can be allocated in one single round is bounded by a constant.

4.1 Framework for Filling-Processes (Caching and Packing)

We make the following assumptions about the processes. For each round $t \geqslant 0$, we sample bin $i = i^t$ and then place a certain number of balls to i (or other bins).

More formally:

Condition $\mathcal{P}_1$: For each round $t = 0, 1, \ldots$, pick an arbitrary labeling of the n bins such that $y_1^t \geqslant y_2^t \geqslant \cdots \geqslant y_n^t$. Then let $i = i^t \in_{p^t} [n]$, where the distribution vector p^t is majorized by the uniform distribution (One-Choice).

Condition $\mathcal{W}_1$: For each round $t = 0, 1, \ldots$, with i being the bin chosen above:

- If $y_i^t < 0$ then allocate exactly $\lceil -y_i^t \rceil + 1 \geqslant 2$ balls such that there can be at most two bins $k_1, k_2 \in [n]$ where:
  (a) $k_1 \in [n]$ receives $\lceil -y_{k_1}^t \rceil + 1$ balls,
  (b) $k_2 \in [n]$ receives $\lceil -y_{k_2}^t \rceil$ balls,
  (c) all other bins $j \in [n]$ receive at most $\lceil -y_j^t \rceil - 1$ balls.
- If $y_i^t \geqslant 0$ then allocate a single ball to bin i.

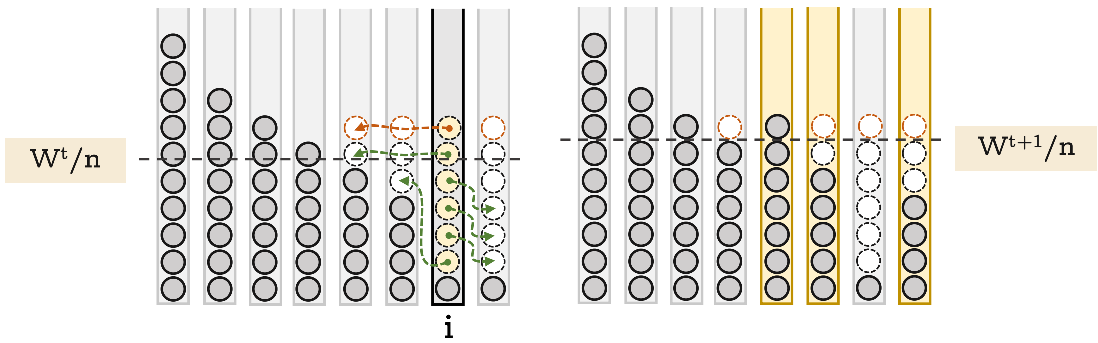

Figure 3: Illustration of an allocation for a filling process. After the underloaded bin i is picked, $\lceil -y_i^t \rceil + 1 = 6$ balls are allocated; 5 of these balls are placed in the underloaded regions of other bins and one ball is placed in the overloaded region (shown in orange).

See Fig. 3 for an example of a process satisfying $\mathcal{W}_1$. A natural example of a process satisfying $\mathcal{P}_1$ and $\mathcal{W}_1$ is Packing, which only allocates to the (initially) chosen bin i.

Lemma 4.1. *The* Packing *process satisfies conditions $\mathcal{P}_1$ and $\mathcal{W}_1$.*

Proof. Note that the Packing process picks a uniform bin i at each round t, thus it satisfies $\mathcal{P}_1$. Furthermore, if $y_i^t < 0$, it allocates exactly $\lceil -y_i^t \rceil + 1$ balls to bin i; otherwise, it allocates one ball to i, and thus $\mathcal{W}_1$ is also satisfied. □

Lemma 4.2. *The* Tight-Packing *process satisfies conditions $\mathcal{P}_1$ and $\mathcal{W}_1$.*

Proof. Note that the Tight-Packing process picks a uniform bin i at each round t, thus it satisfies $\mathcal{P}_1$. Then the allocation satisfies the following properties.

First, in case of $y_i^t < 0$, then: (i) we allocate exactly $\lceil -y_i^t \rceil + 1$ balls, (ii) one bin receives $\lceil -y_j^t \rceil$ balls, and (iii) every other bin j receives a number of balls between $[0, \lceil -y_j^t \rceil - 1]$.

Secondly, in case of $y_i^t \geqslant 0$ we place one ball to i. □

We emphasize that in condition $\mathcal{P}_1$, the distribution p^t may not only be time-dependent, but may also depend on the load distribution at round t, that is, the filtration $\mathfrak{F}^t$. Also, the allocation used in $\mathcal{W}_1$ may depend on the filtration $\mathfrak{F}^t$. Thus the framework also applies in the presence of an adaptive adversary, which directs all the $\lceil -y_i^t \rceil + 1$ balls to be allocated in round t towards the "most loaded" underloaded bins. At the other end of the spectrum, there are natural processes which have a propensity to place these balls into "less loaded" underloaded bins. One specific, more complex example is the CACHING process, where due to the update of the cache after each single ball, the allocation is more skewed towards "less loaded" underloaded bins. However, as we shall discuss shortly, CACHING only satisfies $\mathcal{P}_1$ and $\mathcal{W}_1$ after a suitable compression of rounds.

Next we present our result for the gap of such filling processes:

Theorem 4.3. *Consider any allocation process which satisfies the conditions $\mathcal{P}_1$ and $\mathcal{W}_1$ at each round. Then, there is a constant $C > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant C \log n \,\right] \geqslant 1 - n^{-2}.$$

This result is shown in Section 5 by analyzing an exponential potential function Φ^m over the overloaded bins and eventually establishing that for any $m \geqslant 1$, $\mathbf{E}\left[\, \Phi^m \,\right] = \mathcal{O}(n)$. Then a simple application of Markov's inequality yields the desired result for the gap.

As mentioned above CACHING does not satisfy conditions $\mathcal{P}_1$ and $\mathcal{W}_1$ directly. The issue is that CACHING, as defined in Section 3, allocates only *one* ball at each round whereas, due to condition $\mathcal{W}_1$, a filling process may place several balls into underloaded bins in a single round. To overcome this issue, we define a so-called "**unfolding**" of a filling process. The filling process proceeds in rounds $0, 1, \ldots$, whereas the unfolded version is a coupled process, which is encoded as a sequence of *atomic allocations* (each allocation places exactly one ball into a bin). Using this unfolding along with the gap bound for a filling process satisfying $\mathcal{W}_1$ and $\mathcal{P}_1$, we can then bound the number of atomic allocations with a large gap.

First, we define the process of unfolding formally: Let x^t be the load vector of a filling process P, and $\widehat{x}^t$ be the load vector of the unfolding of P called $U = U(P)$. Initialize $x^1 := \widehat{x}^1$ as the all-zero vector, corresponding to the empty load configuration. For every round $t \geqslant 0$, where P has allocated W^{t-1} balls before, we create atomic allocations $S(t) := \{W^{t-1}+1, \ldots, W^t\}$, where $W^t = W^{t-1} + \min\left\{ \lceil -y_i^t \rceil + 1,\ 1 \right\}$ for process U, such that for each $s \in S(t)$ a single ball is allocated, and we have $x^{t+1} := \widehat{x}^{W^{t+1}}$, i.e., the load distribution of P after round $t+1$ corresponds to load distribution of U after atomic allocation $W^{t+1} \geqslant t+1$. Note that the unfolding of a process is not (completely) unique, as the allocations in $[W^{t-1}+1, W^t]$ of $U(P)$ can be permuted arbitrarily.

Lemma 4.4. *There is a process P satisfying conditions $\mathcal{P}_1$ and $\mathcal{W}_1$, such that for a suitable unfolding $U = U(P)$, the process U is an instance of* CACHING.

Now, applying Theorem 4.3 to the unfolded version of a filling process, and exploiting that in a balanced load configuration not too many atomic allocations can be created through unfolding, we obtain the following result:

Lemma 4.5. *Fix any constant $c > 0$. Then for any filling process P satisfying $\mathcal{W}_1$ and $\mathcal{P}_1$, there is a constant $C = C(c) > 0$ such that for any number of atomic allocations $m \geqslant 1$, with probability at least $1 - n^{-2}$, any unfolding $U = U(P)$ satisfies*

$$\left|\{t \in [m] \colon\ \mathrm{Gap}_U(t) \leqslant C \cdot \log n\}\right| \geqslant n^{-c} \cdot m \log m.$$

Hence for any m, all but a polynomially small fraction of the first m atomic allocations in any unfolding of a filling process (e.g. CACHING) have a logarithmic gap. This behavior matches the one of the original filling process; the only limitation is that we cannot prove a small gap that holds for an arbitrarily large *fixed* atomic allocation m. However, if m is polynomial in n, that is, $m \leqslant n^c$ for some constant $c > 0$, then Lemma 4.5 implies directly that with high probability, the gap at atomic allocation m (and at all atomic allocations before) is logarithmic.

Returning to the special case of CACHING, we will now perform a more tailored analysis to obtain the gap bound analogous to Theorem 4.3 so that it holds for an arbitrarily large number of atomic allocations of the unfolding:

Lemma 4.6. *For the* CACHING *process (which allocates exactly one ball per round) there is a constant $C > 0$ such that for any number of rounds $m \geqslant 1$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant C \cdot \log n \,\right] \geqslant 1 - n^{-2}.$$

4.2 Framework for Non-Filling-Processes (Thinning and Twinning)

Recall that for any $i \in [n]$, p_i^t is the probability of picking the i-th most heavily loaded bin for an allocation at round t. As before, p_i^t may be chosen *adaptively*, i.e., dependent on the load vector x^t (equivalently, based on the filtration $\mathfrak{F}^t$).

Our allocation process for non-filling work as follows. First, we define $p_+^t := \max_{i \in B_+^t} p_i^t$ and $p_-^t := \min_{i \in B_-^t} p_i^t$, as the largest (smallest) probability for allocating to a fixed overloaded (underloaded) bin at round t, respectively. As in filling processes, we then first sample a bin $i = i^t \in [n]$ and then place a certain number of balls to i. More formally,

Condition $\mathcal{P}_2$: For each round $t = 0, 1, \ldots$, we consider an arbitrary labeling such that $y_1^t \geqslant y_2^t \geqslant \cdots \geqslant y_n^t$. Then let $i = i^t \in_{p^t} \in [n]$, where the distribution vector p^t must satisfy $p_+^t \leqslant \frac{1}{n} \leqslant p_-^t$.

Condition $\mathcal{W}_2$: For each round $t = 0, 1, \ldots$, when i is chosen for allocation,

– If $y_i^t < 0$, then place w_- balls into bin i,

– If $y_i^t \geqslant 0$, then place w_+ balls into bin i,

where $1 \leqslant w_+ \leqslant w_-$ are constant integers independent of t and n.

Both conditions are natural, but on their own they are not sufficient to establish a good bound on the gap, as the ONE-CHOICE process satisfies both conditions with equality. Thus, we will require processes to satisfy at least one of two stronger versions of $\mathcal{P}_2$ and $\mathcal{W}_2$:

Condition $\mathcal{P}_3$: This is as Condition $\mathcal{P}_2$, but additionally, there are time-independent constants $k_1 \in (0, 1], k_2 \in (0, 1]$ such that for each round $t \geqslant 0$:

$$p_+^t \leqslant \frac{1 - k_1}{n} + \frac{k_1 \cdot |B_+^t|}{n^2} = \frac{1}{n} - \frac{k_1 \cdot (1 - \delta^t)}{n},$$
$$p_-^t \geqslant \frac{1}{n} + \frac{k_2 \cdot |B_+^t|}{n^2} = \frac{1}{n} + \frac{k_2 \cdot \delta^t}{n}.$$

Condition $\mathcal{W}_3$: This is as Condition $\mathcal{W}_2$, but additionally we have the strict inequality: $w_+ < w_-$. Also, we assume that for each $t \geqslant 0$, distribution vector p_i^t is non-decreasing in i.

The reason we attach the non-decreasing property of p^t to $\mathcal{W}_3$ and not to $\mathcal{P}_3$ is to make our main result slightly stronger.

The rationale behind condition $\mathcal{P}_3$ is that we wish to slightly bias the distribution vector p^t towards underloaded bins at each round t. However, it is natural to assume that this influence is limited by a process that samples, say, at most two bins uniformly and independently, and then allocates balls to the least loaded of the two. Concretely, if a process takes *two* independent and uniform bin samples at each round, the probability of picking two overloaded bins equals $\big(\frac{|B_+^t|}{n}\big)^2$. Hence by averaging, there must be a bin $i \in B_+^t$ such that

$$p_+^t \geqslant p_i^t \geqslant \left(\frac{|B_+^t|}{n}\right)^2 \cdot \frac{1}{|B_+^t|} = \frac{|B_+^t|}{n^2}.$$

The relaxation of the first constraint in $\mathcal{P}_3$ by taking a strict convex combination of $\frac{1}{n}$ and $\frac{|B_+^t|}{n^2}$ ensures some slack, for instance, it allows the framework to cover a "noisy" version of Mean-Thinning, where at each round, we perform a One-Choice allocation with some constant probability < 1, and otherwise we perform an allocating following the Mean-Thinning process (see Lemma 4.11 for details). Similarly, for any process which takes at most two uniform samples, by averaging, there must be a bin $j \in B_-^t$ such that

$$p_-^t \leqslant p_j^t \leqslant \frac{1 - \frac{|B_+^t|^2}{n^2}}{|B_-^t|} = \frac{(n - |B_+^t|) \cdot (n + |B_+^t|)}{n^2 |B_-^t|} = \frac{1}{n} + \frac{|B_+^t|}{n^2}.$$

Finally, we remark that $\mathcal{P}_3$ resembles the framework of [24, Equation 2], where $p_i^t = p_i$ is non-decreasing in i and $p_{n/3} \leqslant \frac{1-4\varepsilon}{n}$ and $p_{2n/3} \geqslant \frac{1+4\varepsilon}{n}$ holds for some $0 < \varepsilon < 1/4$ (not necessarily constant). In contrast to that, for constants $k_1, k_2 > 0$, the conditions in $\mathcal{P}_3$ are relaxed as they only imply such a bias if δ^t is bounded away from 0 and 1, which may not hold in all rounds.

Lemma 4.7. *The* Twinning *process satisfies $\mathcal{P}_2$ and $\mathcal{W}_3$.*

Proof. The statement for the Twinning process is obvious, since $w_- = 2 > 1 = w_+$ and the bin i is picked uniformly at each round. □

Lemma 4.8. *The* Mean-Thinning *process satisfies $\mathcal{P}_3$ and $\mathcal{W}_2$.*

Proof. For the Mean-Thinning process, we use the following equivalent two-stage description: Each ball is allocated to the first sample, if the bin is underloaded; otherwise, the ball is allocated to the second sample. Hence the probability of allocating to any overloaded bin $i \in B_+^t$ is $p_i^t = \frac{\delta^t}{n} = \frac{1}{n} - \frac{1 \cdot (1-\delta^t)}{n}$, so we can choose $k_1 := 1$. For any underloaded bin $i \in B_-^t$, $p_i^t = \frac{1+\delta^t}{n} = \frac{1}{n} + \frac{1 \cdot \delta^t}{n}$, so we can choose $k_2 := 1$, and $\mathcal{P}_3$ holds. □

By an inductive argument involving adding "extra balls", we can reduce Thinning$(f(n))$ with a non-negative threshold function $f(n)$, possibly depending on n but independent of t, to Mean-Thinning. Note that Thinning with a suitable positive function $f(n)$ is appealing, as results from [10] have shown a sublogarithmic gap in the lightly loaded case where $m = \mathcal{O}(n)$ (see also Table 1).

Lemma 4.9. *Let $f(n)$ be any non-negative function. Let Gap_0 and $\mathrm{Gap}_{f(n)}$ be the gaps of* Mean-Thinning *and* Thinning$(f(n))$. *Then $\mathrm{Gap}_{f(n)}$ is stochastically smaller than $\mathrm{Gap}_0 + f(n)$.*

Before proving the lemma, we need the following domination result:

Lemma 4.10. *Let process P_A be a* Thinning *process with an empty load distribution at round 0, and using threshold $t/n + f(n)$ at any round $t \geqslant 0$, where $f(n)$ is non-negative. Further, let process P_B be a* Thinning *process with initial load distribution $x_1^0 = x_2^0 = \cdots = x_n^0 = f(n)$, and using threshold $t/n + f(n)$ at any round $t \geqslant 0$. Then, there is a coupling so that at any round $t \geqslant 0$ and for any bin $i \in [n]$, $x_i^t(A) \leqslant x_i^t(B)$.*

Proof of Lemma 4.10. For any round $t \geqslant 0$, let $i_1 = i_1^t$ and $i_2 = i_2^t$ the two random bin samples, which are uniform and independent over $[n]$. We consider a coupling between P_A and P_B, where these random bin samples are identical, and prove inductively that for any $t \geqslant 0$ and any $i \in [n]$,

$$x_i^t(A) \leqslant x_i^t(B).$$

The base case $t = 0$ holds by definition. For the induction step, we make a case distinction:
Case 1 $[x_{i_1}^t(A) < t/n + f(n)]$. In this instance P_A allocates a ball to i_1. If $x_{i_1}^t(B) < t/n + f(n)$, then P_B also allocates a ball to i_2; otherwise, we have $x_{i_1}^t(B) \geqslant t/n + f(n)$, and hence $x_{i_1}^t(B) > x_{i_1}^t(A)$, i.e., $x_{i_1}^t(B) \geqslant x_{i_1}^t(A) + 1$. This implies

$$x_{i_1}^{t+1}(B) = x_{i_1}^t(B) \geqslant x_{i_1}^t(A) + 1 = x_{i_1}^{t+1}(A),$$

and the inductive step follows from this and the induction hypothesis.

Case 2 $[x_{i_1}^t(A) \geqslant t/n + f(n)]$. In this instance P_A allocates a ball to i_2. By induction hypothesis, $x_{i_1}^t(A) \leqslant x_{i_1}^t(B)$, which implies P_B also allocates a ball to i_2. Thus we have

$$x_{i_1}^{t+1}(A) = x_{i_1}^t(A) + 1 \geqslant x_{i_1}^t(B) + 1 = x_{i_1}^{t+1}(B),$$

and the inductive step is complete.

Since in both cases all other bins remain unchanged the proof is complete. □

We can now complete the proof of Lemma 4.9.

Proof of Lemma 4.9. Instead of starting with an empty load distribution, we modify it by adding $f(n)$ extra balls to each bin. We refer to these as the *red* balls. We now run the Thinning process with threshold $(t/n + f(n))$ on these bins, assigning one new *blue* ball each round. Due to the presence of the $f(n)$ red balls in each bin the process will consider any bin with more than t/n blue balls as being overloaded. Thus from the perspective of the blue balls this is a Mean-Thinning process.

Using the coupling from Lemma 4.10, adding $f(n) \cdot n$ red balls in total without changing the threshold, always results in a load vector that point-wise majorizes the unmodified Thinning process with threshold $t/n + f(n)$. □

Let us now define a noisy version of Mean-Thinning, called $(1 + \eta)$-Mean-Thinning, where $\eta \in (0, 1]$ is a parameter, analogous to β in the $(1 + \beta)$-process. At each round $t \geqslant 0$, with probability η, we allocate a ball using Mean-Thinning; otherwise we allocate a ball using One-Choice.

Lemma 4.11. *For any constant $\eta > 0$, the $(1 + \eta)$-Mean-Thinning process satisfies $\mathcal{P}_3$ and $\mathcal{W}_2$.*

Proof. Let p be the distribution vector of $(1 + \eta)$-Mean-Thinning with parameter $\eta \in (0, 1)$; that is, with probability η it performs Mean-Thinning, otherwise One-Choice. Then for any $i \in B_+^t$,

$$p_i^t = (1 - \eta) \cdot \frac{1}{n} + \eta \cdot \frac{\delta^t}{n} = \frac{1}{n} - \frac{\eta \cdot (1 - \delta^t)}{n}.$$

Similarly, for any $i \in B_-^t$,

$$p_i^t = (1 - \eta) \cdot \frac{1}{n} + \eta \cdot \frac{1 + \delta^t}{n} = \frac{1}{n} + \frac{\eta \cdot \delta^t}{n}.$$

Thus for $k_1 = k_2 = \eta \in (0, 1)$, the $(1 + \eta)$-Mean-Thinning process satisfies $\mathcal{P}_3$ and $\mathcal{W}_2$. □

While the $(1+\beta)$-process does not satisfy $\mathcal{P}_3$ at each round, the process can be majorized by $(1+\eta)$ with $\eta = \beta$.

Lemma 4.12. *For any constant $\beta \in (0,1]$, let Gap_β be the gap of the $(1+\beta)$-process. Then $(1+\eta)$ process with $\eta = \beta$ majorizes $(1+\beta)$ at each round. Hence, Gap_β is stochastically smaller than Gap_η, the gap of the $(1+\eta)$ process.*

Before proving this lemma, we state one auxiliary result, which is implicit in [24]. Theorem 3.1 proves the required majorization for time-independent distribution vectors, but as mentioned in the proof of Theorem 3.2 [24], the same result generalizes to time-dependent distribution vectors.

Lemma 4.13 (see [24, Section 3]). *Consider two allocation processes Q and P with $w_+ = w_- = 1$. The allocation process Q uses at each round a fixed allocation distribution q. The allocation process P uses a time-dependent allocation distribution p^t, which may depend on $\mathfrak{F}^t$ but majorizes q at each round $t \geqslant 0$. Let $y^t(Q)$ and $y^t(P)$ be the two load vectors, sorted decreasingly. Then there is a coupling such that for all rounds $t \geqslant 0$, $y^t(Q)$ is majorized by $y^t(P)$.*

Proof of Lemma 4.12. We will show that for any round $t \geqslant 0$ and for any load configuration, the $(1+\eta)$ for $\eta = \beta$ distribution vector majorizes the distribution vector of $(1+\beta)$-process. So, by Lemma 4.13, the claim will follow.

Recall that the $(1+\beta)$ distribution vector q^t is given by,

$$q_i^t = q_i = (1-\beta) \cdot \frac{1}{n} + \frac{\beta(2i-1)}{n^2}.$$

The distribution vector p^t for the $(1+\eta)$ process is non-decreasing and uniform over B_-^t and B_+^t, so majorization follows immediately once we prove that

$$\sum_{i=1}^{|B_+^t|} q_i \leqslant \sum_{i=1}^{|B_+^t|} p_i^t.$$

For the distribution vector q we have,

$$\sum_{i=1}^{\delta^t \cdot n} q_i = \sum_{i=1}^{\delta^t \cdot n} (1-\beta) \cdot \frac{1}{n} + \sum_{i=1}^{\delta^t \cdot n} \frac{\beta(2i-1)}{n^2} = (1-\beta) \cdot \delta^t + \frac{\beta(\delta^t \cdot n)^2}{n^2} = \delta^t - \beta \cdot (\delta^t - (\delta^t)^2).$$

Similarly, for the distribution vector p^t we have,

$$\sum_{i=1}^{\delta^t \cdot n} p_i^t = \sum_{i=1}^{\delta^t \cdot n} \frac{1}{n} - \sum_{i=1}^{\delta^t \cdot n} \frac{\beta(1-\delta^t)}{n} = \delta^t - \beta \cdot (\delta^t - (\delta^t)^2). \qquad \square$$

Since $\beta = 1$ yields Two-Choice, our framework also applies to the Two-Choice process.

The following basic, yet crucial result follows from the preconditions in Theorem 4.15:

Lemma 4.14. *Consider any process satisfying the conditions $\mathcal{P}_3$ and $\mathcal{W}_2$, or, $\mathcal{P}_2$ and $\mathcal{W}_3$. Then for the constant $c_1 := c_1(k_1, k_2) > 0$, it holds that for any $t \geqslant 0$,*

$$p_-^t \cdot w_- - p_+^t \cdot w_+ \geqslant \frac{c_1}{n}.$$

Proof. First assume $\mathcal{P}_2$ and $\mathcal{W}_3$ holds. In this case, $\mathcal{P}_2$ implies $p_-^t \geqslant \frac{1}{n} \geqslant p_+^t$, and thus

$$p_-^t \cdot w_- - p_+^t \cdot w_+ \geqslant \frac{1}{n} \cdot w_- - \frac{1}{n} \cdot w_+ \geqslant \frac{c_1}{n},$$

for $c_1 := w_- - w_+ > 0$, since the weights are constants satisfying $w_- > w_+$.

Next assume $\mathcal{P}_3$ and $\mathcal{W}_2$ holds. In this case, $\mathcal{W}_2$ implies $w_- \geqslant w_+ \geqslant 1$. Using $\mathcal{P}_2$,

$$\begin{aligned} p_-^t \cdot w_- - p_+^t \cdot w_+ &\geqslant p_-^t \cdot w_+ - p_+^t \cdot w_+ \geqslant (p_-^t - p_+^t) \cdot 1 \\ &\geqslant \Big(\frac{1}{n} + \frac{k_2 \cdot \delta^t}{n} - \frac{1}{n} + \frac{k_1 \cdot (1-\delta^t)}{n}\Big) \geqslant \frac{c_1}{n}, \end{aligned}$$

for $c_1 := \min\{k_1, k_2\}$. $\square$

It is important to highlight that the quantity $p_-^t \cdot w_- - p_+^t \cdot w_+$ does not involve the *number* of underloaded/underloaded bins. Hence even if we know that $p_-^t \cdot w_- - p_+^t \cdot w_+ > 0$ holds, then this does not necessarily imply that the expected weight allocated to the set of underloaded bins is larger than the expected weight allocated to the set of overloaded bins. This conclusion would only be true if the quantile δ^t is sufficiently close to $1/2$.

One large portion of the analysis is devoted to derive a *weaker* quantile condition, that is, we prove that for sufficiently many rounds, the quantile δ^t is in the range $(\varepsilon, 1-\varepsilon)$ for some (small) constant ε. For this analysis, the inequality in Lemma 4.14 will be useful when we establish a connection between the absolute and quadratic potential function (Lemma 6.2).

Theorem 4.15. *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_3 \cap \mathcal{W}_3$-process, there exists a constant $\kappa > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\max_{i\in[n]}\left|x_i^m - \frac{W^t}{n}\right| \leqslant \kappa \log n\right] \geqslant 1 - n^{-3};$$

so in particular, $\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \log n\,\right] \geqslant 1 - n^{-3}$.

Thanks to the reductions in Lemma 4.9 and Lemma 4.12, we also deduce:

Corollary 4.16. *For the $(1+\beta)$-process for any constant $\beta \in (0,1]$ and for any* Thinning *$(f(n))$-process where $f(n) \in [0, \mathcal{O}(\log n)]$ there exists a constant $\kappa > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \log n\,\right] \geqslant 1 - n^{-3}.$$

5 Analysis of Filling Processes

In this section, we present our analysis for filling processes. We begin in Section 5.1 by defining the exponential potential function and analyze its behavior (depending on some other constraints). Then in Section 5.2 we use a super-martingale argument to show that the exponential potential function decreases, which eventually yields the desired gap bound in Theorem 4.3. In Section 5.3 we present our results on (general) unfoldings of filling processes. We also prove that Caching can be seen as the unfolded version of a filling process satisfying $\mathcal{W}_1$ and $\mathcal{P}_1$, and thus deduce $\mathcal{O}(\log n)$ gap bound as long as the number of rounds is polynomial in n. Finally, in Section 5.4 we strengthen this by showing an $\mathcal{O}(\log n)$ upper bound on the gap for the Caching process which holds for any $m \geqslant 1$.

5.1 Potential Function Analysis

We consider a version of the exponential potential function Φ^t which only takes bins into account whose load is at least two above the average load. This is given by

$$\Phi^t := \sum_{i:y_i^t \geqslant 2} \exp\left(\alpha \cdot y_i^t\right) = \sum_{i=1}^{n} \exp\left(\alpha \cdot y_i^t\right) \cdot \mathbf{1}_{\{y_i^t \geqslant 2\}},$$

where we recall that $y_i^t = x^t - \frac{W^t}{n}$ is the normalized load of bin i at round t and $\alpha > 0$ is a sufficiently small constant to be fixed later. Let $\Phi_i^t = \exp\left(\alpha \cdot y_i^t\right) \cdot \mathbf{1}_{\{y_i^t \geqslant 2\}}$ and thus $\Phi^t = \sum_{i=1}^n \Phi_i^t$. We will also use the absolute value potential:

$$\Delta^t := \sum_{i=1}^{n} \left|y_i^t\right|.$$

The next lemma provides a useful upper bound on the expected potential. It establishes that to bound $\mathbf{E}[\,\Phi^{t+1} \mid \mathfrak{F}^t\,]$ from above, we may assume the distribution vector p^t is uniform.

Lemma 5.1. *Consider any allocation process satisfying $\mathcal{P}_1$ and $\mathcal{W}_1$. Then for any round $t \geqslant 0$,*

$$\mathbf{E}[\,\Phi^{t+1} \mid \mathfrak{F}^t\,] \leqslant \frac{1}{n} \sum_{i=1}^{n} \Phi_i^t \cdot \left(\sum_{j\,:\,y_j^t < 1} e^{\frac{-\alpha(\lceil -y_j^t \rceil + 1)}{n}} + e^{-\frac{\alpha}{n}} \cdot (|B_{\geqslant 1}^t| - 1) + e^{\alpha - \frac{\alpha}{n}} \right) + e^{3\alpha},$$

where $B_{\geqslant 1}^t$ denotes the set of bins with load at least 1*.*

Proof. Recall that the filtration $\mathfrak{F}^t$ reveals the load vector x^t. Throughout this proof, we consider the labeling chosen by the process in round t such that $x_1^t \geqslant x_2^t \geqslant \cdots \geqslant x_n^t$ and p^t being majorized by ONE-CHOICE (according to $\mathcal{P}_1$). We emphasize that for this labeling, x_i^{t+1} may not be non-increasing in $i \in [n]$.

To begin, using $\Phi_i^{t+1} = e^{\alpha y_i^{t+1}} \mathbf{1}_{\{y_i^{t+1} \geqslant 2\}}$, we can split Φ^{t+1} over the n bins as follows,

$$\mathbf{E}\left[\,\Phi^{t+1} \mid \mathfrak{F}^t\,\right] = \sum_{j=1}^{n} \mathbf{E}\left[\,\Phi_j^{t+1} \mid \mathfrak{F}^t\,\right].$$

Now consider the effect of picking bin i for the allocation in round t to the potential Φ^{t+1}. Note that bin i is chosen with probability equal to p_i^t. Observe that if bin i satisfies $y_i^t < 1$, then it receives at most $\lceil -y_i^t \rceil + 1$ balls in round t, thus $y_i^{t+1} \geqslant 2$ if and only if $y_i^t \geqslant 1$. Using condition $\mathcal{W}_1$, we distinguish between the following three cases based on how allocating to i changes Φ_j^t for $j \neq i$ and for $j = i$:

Case 1.A $[y_i^t < 1,\, j \neq i,\, y_j^t < 1]$. We will allocate $\lceil -y_i^t \rceil + 1$ many balls to bins k with $y_k^t < 1$ (not necessarily to i) subject to $\mathcal{W}_1$. This increases the average load by $(\lceil -y_i^t \rceil + 1)/n$. Since $y_j^t < 1$, $\Phi_j^t = 0$. Further, by condition $\mathcal{W}_1$ we can increase the load of y_j^t by at most $\lceil -y_j^t \rceil + 1$, hence, $\Phi_j^{t+1} = 0$. Therefore, $\Phi_j^{t+1} = \Phi_j^t \cdot e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}}$ for $j \neq i$.

Case 1.B $[y_i^t < 1,\, j \neq i,\, y_j^t \geqslant 1]$. As in Case 1a, we will allocate $\lceil -y_i^t \rceil + 1$ many balls to bins k with $y_k^t < 1$ (not necessarily to i) subject to $\mathcal{W}_1$, which increases the average load by $(\lceil -y_i^t \rceil + 1)/n$. Additionally, bin j will receive no balls (by condition $\mathcal{W}_1$), thus $\Phi_j^{t+1} = \Phi_j^t \cdot e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}}$ for $j \neq i$.

Case 2 $[y_i^t \geqslant 1,\, j \neq i]$. We allocate one ball to i, which increases the average load by $1/n$, and thus $\Phi_j^{t+1} = \Phi_j^t \cdot e^{-\frac{\alpha}{n}}$ for $j \neq i$, which again also holds for bins j that do not contribute.

Case 3 [$j = i$]. Finally, we consider the effect on Φ_i^{t+1}. Again if $y_i^t < 1$, then $\Phi_i^t = 0$ and $\Phi_i^{t+1} = 0$. Otherwise, we have $y_i^t \geqslant 1$ and we allocate one ball to i, and thus

$$\Phi_i^{t+1} = e^{\alpha \cdot (y_i^t+1) - \frac{\alpha}{n}} \mathbf{1}_{\{y_i^{t+1} \geqslant 2\}} = e^{\alpha \cdot y_i^t} \mathbf{1}_{\{y_i^{t+1} \geqslant 2\}} \cdot e^{\alpha - \frac{\alpha}{n}} \leqslant (e^{\alpha \cdot y_i^t} \mathbf{1}_{\{y_i^t \geqslant 2\}} + e^{2\alpha}) \cdot e^{\alpha - \frac{\alpha}{n}},$$

where the $+e^{2\alpha}$ is added to account for the case where $1 \leqslant y_i^t < 2$ and so $\Phi_i^t = 0$ but $0 < \Phi_i^{t+1} \leqslant e^{3\alpha - \alpha/n}$. Thus in this case, $\Phi_i^{t+1} \leqslant \Phi_i^t e^{\alpha - \frac{\alpha}{n}} + e^{3\alpha - \frac{\alpha}{n}} \leqslant \Phi_i^t e^{\alpha - \frac{\alpha}{n}} + e^{3\alpha}$.

By aggregating the three cases above, and observing that $\sum_{i=1}^n p_i^t \cdot e^{3\alpha} = e^{3\alpha}$, we see that

$$\begin{aligned}
\sum_{j=1}^n \mathbf{E}[\, \Phi_j^{t+1} \mid \mathfrak{F}^t \,] &= \sum_{i=1}^n p_i^t \sum_{j=1}^n \mathbf{E}[\, \Phi_j^{t+1} \mid \mathfrak{F}^t, \text{Bin i is selected at round t} \,] \\
&\leqslant \sum_{i=1}^n p_i^t \cdot \mathbf{1}_{\{y_i^t < 1\}} \sum_{j=1}^n \Phi_j^t \cdot e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}} + \sum_{i=1}^n p_i^t \cdot \mathbf{1}_{\{y_i^t \geqslant 1\}} \sum_{j \neq i} \Phi_j^t \cdot e^{-\frac{\alpha}{n}} \\
&\quad + \sum_{i=1}^n p_i^t \cdot \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot \Phi_i^t \cdot e^{\alpha - \frac{\alpha}{n}} + e^{3\alpha} \qquad (5.1)
\end{aligned}$$

We will now rewrite (5.1) in order to establish that it is maximized if p is the uniform distribution. Adding $\sum_{i=1}^n p_i^t \cdot \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot \Phi_i^t \cdot e^{-\frac{\alpha}{n}}$ to the middle sum (corresponding to Case 2) in (5.1) and subtracting it from the last sum (corresponding to Case 3) transforms (5.1) into

$$\begin{aligned}
&\sum_{i=1}^n p_i^t \mathbf{1}_{\{y_i^t < 1\}} \sum_{j=1}^n \Phi_j^t e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}} + \sum_{i=1}^n p_i^t \mathbf{1}_{\{y_i^t \geqslant 1\}} \sum_{j=1}^n \Phi_j^t e^{-\frac{\alpha}{n}} + \sum_{i=1}^n p_i^t \Phi_i^t \mathbf{1}_{\{y_i^t \geqslant 1\}} e^{-\frac{\alpha}{n}} (e^{\alpha} - 1) + e^{3\alpha} \\
&= \sum_{i=1}^n p_i^t \left(\Bigg(\underbrace{\mathbf{1}_{\{y_i^t < 1\}} e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}} + \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot e^{-\frac{\alpha}{n}}}_{g(i)} \Bigg) \sum_{j=1}^n \Phi_j^t + \underbrace{\Phi_i^t \cdot \mathbf{1}_{\{y_i^t \geqslant 1\}} e^{-\frac{\alpha}{n}} (e^{\alpha} - 1)}_{f(i)} \right) + e^{3\alpha}
\end{aligned}$$

Recall that $y_1^t \geqslant y_2^t \geqslant \cdots \geqslant y_n^t$, which implies $\Phi_1^t \geqslant \Phi_2^t \geqslant \cdots \geqslant \Phi_n^t \geqslant 0$. Thus $f(i)$ and $g(i)$ are non-negative and non-increasing in i and $\sum_{j=1}^n \Phi_j^t \geqslant 0$. Consequently, the function $h(i) = f(i) + g(i) \cdot \sum_{j=1}^n \Phi_j^t$ is non-negative and non-increasing in i. Note that by condition $\mathcal{P}_1$, for any $k \in [n]$ it holds that $\sum_{i=1}^k p_i^t \leqslant \frac{k}{n}$. Thus we can apply Lemma A.2 which implies $\sum_{i=1}^n p_i^t \cdot h(i) \leqslant \sum_{i=1}^n \frac{1}{n} \cdot h(i)$. Applying this to the above, rearranging, and splitting $f(i)$ gives

$$\begin{aligned}
\mathbf{E}[\, \Phi^{t+1} \mid \mathfrak{F}^t \,] \leqslant \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t < 1\}} e^{\frac{-\alpha(\lceil -y_i^t \rceil + 1)}{n}} \sum_{j=1}^n \Phi_j^t + \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot e^{-\frac{\alpha}{n}} \sum_{j=1}^n \Phi_j^t \\
- \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \Phi_i^t \cdot e^{-\frac{\alpha}{n}} + \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \Phi_i^t \cdot e^{\alpha - \frac{\alpha}{n}} + e^{3\alpha}
\end{aligned} \qquad (5.2)$$

Now observe combining the second and third terms above gives

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot e^{-\frac{\alpha}{n}} \sum_{j=1}^n \Phi_j^t - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \Phi_i^t \cdot e^{-\frac{\alpha}{n}} &= \frac{1}{n} \cdot e^{-\frac{\alpha}{n}} \cdot \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{y_i^t \geqslant 1\}} \cdot \Phi_j^t \cdot \mathbf{1}_{\{j \neq i\}} \\
&= \frac{1}{n} \cdot e^{-\frac{\alpha}{n}} \cdot \sum_{i=1}^n \Phi_i^t \sum_{j=1}^n \mathbf{1}_{\{y_j^t \geqslant 1\}} \cdot \mathbf{1}_{\{j \neq i\}} \\
&= \frac{1}{n} \cdot e^{-\frac{\alpha}{n}} \cdot \sum_{i=1}^n \Phi_i^t \cdot (|B_{\geqslant 1}^t| - 1), \qquad (5.3)
\end{aligned}$$

where the last line follows since $\Phi_i^t = e^{\alpha y_i^t}\mathbf{1}_{\{y_i^t \geqslant 2\}}$.

Now, substituting (5.3) into (5.2), exchanging the first double summation and using the bound $\mathbf{1}_{\{y_i^t \geqslant 1\}} \leqslant 1$ on the last sum, and finally grouping terms gives the following

$$\mathbf{E}[\,\Phi^{t+1} \mid \mathfrak{F}^t\,] \leqslant \frac{1}{n}\sum_{i=1}^{n}\Phi_i^t\left(\sum_{j=1}^{n}\mathbf{1}_{\{y_j^t<1\}}e^{\frac{-\alpha(\lceil -y_j^t\rceil+1)}{n}} + (|B_{\geqslant 1}^t|-1)e^{-\frac{\alpha}{n}} + e^{\alpha-\frac{\alpha}{n}}\right) + e^{3\alpha},$$

as claimed. $\square$

Let $\mathcal{G}^t$ be the event that at round $t \geqslant 0$ either there are at least $n/20$ underloaded bins or there is a an absolute value potential of at least $n/10$. In symbols this is given by

$$\mathcal{G}^t := \left\{B_-^t \geqslant n/20\right\} \cup \left\{\Delta^t \geqslant n/10\right\}. \tag{5.4}$$

The next lemma provides two estimates on the expected exponential potential Φ^{t+1} in terms of Φ^t. The first estimate holds for any round and it establishes that the process does not perform worse than One-Choice, meaning that the potential increases by a factor of at most $(1 + \mathcal{O}(\alpha^2/n))$. The second estimate is stronger for rounds where we have a lot of underloaded bins or a large value of the absolute value potential. This stronger estimate states that the potential decreases by a factor of $(1 - \Omega(\alpha/n))$ in those rounds. Note that as we show in Claim B.1, the potential may increase in expectation for certain load configurations, so it seems hard to prove a decrease without additional constraints.

Lemma 5.2. *Consider any allocation process satisfying $\mathcal{P}_1$ and $\mathcal{W}_1$. There exists a constant $c_1 > 0$ such that for any $0 < \alpha < 1$ and any $t \geqslant 0$,*

$$\mathbf{E}\left[\,\Phi^{t+1} \mid \mathfrak{F}^t\,\right] \leqslant \left(1 + \frac{c_1\alpha^2}{n}\right) \cdot \Phi^t + e^{3\alpha}.$$

Further, there exists a constant $c_2 > 0$ such that for any $0 < \alpha < 1/100$ and any $t \geqslant 0$,

$$\mathbf{E}\left[\,\Phi^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t\,\right] \leqslant \left(1 - \frac{c_2\alpha}{n}\right) \cdot \Phi^t + e^{3\alpha}.$$

Proof. Recall that, as before, we fix the labeling chosen by the process in round t such that $x_1^t \geqslant x_2^t \geqslant \cdots \geqslant x_n^t$, thus x_i^{t+1} may not be non-increasing in $i \in [n]$ and p^t being majorized by One-Choice (according to $\mathcal{P}_1$).

Let A_i be the bracketed term in the expression for $\mathbf{E}[\,\Phi^{t+1} \mid \mathfrak{F}^t\,]$ in Lemma 5.2, given by

$$A_i = \sum_{j\,:\, y_j^t<1} e^{\frac{-\alpha(\lceil -y_j^t\rceil+1)}{n}} + (|B_{\geqslant 1}^t|-1)\cdot e^{-\frac{\alpha}{n}} + e^{\alpha-\frac{\alpha}{n}}. \tag{5.5}$$

Observe that $-\alpha(\lceil -y_j^t\rceil + 1) \leqslant -\alpha$ whenever $y_j^t < 1$ and thus

$$A_i \leqslant \sum_{j\,:\, y_j^t<1} e^{-\frac{\alpha}{n}} + (|B_{\geqslant 1}^t|-1)\cdot e^{-\frac{\alpha}{n}} + e^{\alpha-\frac{\alpha}{n}} = e^{-\frac{\alpha}{n}}\cdot\left[n-1+e^{\alpha}\right].$$

Applying the Taylor estimate $e^z \leqslant 1 + z + z^2$, which holds for any $z \leqslant 1$, twice gives

$$A_i \leqslant \left(1 - \frac{\alpha}{n} + \frac{\alpha^2}{n^2}\right)\left(n+\alpha+\alpha^2\right) = n\cdot\left(1 - \frac{\alpha}{n} + \frac{\alpha^2}{n^2}\right)\left(1 + \frac{\alpha}{n} + \frac{\alpha^2}{n}\right) \leqslant n\cdot\left(1 + \frac{c_1\alpha^2}{n}\right),$$

for some constant $c_1 > 0$. The first statement in the lemma now follows as Lemma 5.1 gives

$$\mathbf{E}[\,\Phi^{t+1} \mid \mathfrak{F}^t\,] \leqslant \frac{1}{n}\sum_{i=1}^{n}\Phi_i^t\cdot A_i + e^{3\alpha} \leqslant \frac{1}{n}\cdot n\left(1 + \frac{c_1\alpha^2}{n}\right)\cdot\sum_{i=1}^{n}\Phi_i^t + e^{3\alpha} \leqslant \left(1 + \frac{c_1\alpha^2}{n}\right)\Phi^t + e^{3\alpha}.$$

We shall now show the second statement of the lemma. By splitting sums in (5.5) we have

$$
\begin{aligned}
A_i &= \sum_{j \in B_-^t} e^{\frac{-\alpha(\lceil -y_j^t \rceil + 1)}{n}} + \sum_{j \colon 0 \leqslant y_j^t < 1} e^{\frac{-\alpha(\lceil -y_j^t \rceil + 1)}{n}} + (|B_{\geqslant 1}^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{\alpha - \frac{\alpha}{n}} \\
&= \sum_{j \in B_-^t} e^{\frac{-\alpha(\lceil -y_j^t \rceil + 1)}{n}} + (|B_+^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{\alpha - \frac{\alpha}{n}} \qquad (5.6) \\
&\leqslant |B_-^t| \cdot e^{-\frac{2\alpha}{n}} + (|B_+^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{\alpha - \frac{\alpha}{n}} \qquad (5.7)
\end{aligned}
$$

We shall now first assume that $|B_-^t| \geqslant n/20$. Recall the bound $e^x \leqslant 1 + x + 0.6 \cdot x^2$ which holds for any $x \leqslant 1/2$. If $\alpha < 1/2$, we can apply this bound to (5.7), giving

$$
\begin{aligned}
A_i &\leqslant e^{-\frac{\alpha}{n}} \cdot \left(|B_-^t| \cdot \left(1 - \frac{\alpha}{n} + \frac{6\alpha^2}{10n^2}\right) + (|B_+^t| - 1) + 1 + \alpha + \frac{6\alpha^2}{10} \right) \\
&\leqslant \left(1 - \frac{\alpha}{n} + \frac{6\alpha^2}{10n^2}\right) \cdot n \left(1 - \frac{\alpha}{20n} + \frac{6\alpha^2}{200n^2} + \frac{\alpha}{n} + \frac{6\alpha^2}{10n}\right) \\
&= n \cdot \left(1 - \frac{\alpha(1 - 12\alpha)}{20n} + \mathcal{O}\left(\frac{\alpha^2}{n^2}\right)\right). \qquad (5.8)
\end{aligned}
$$

We now assume that $|\Delta^t| \geqslant n/10$. Observe that by Schur-convexity (see Lemma A.5) and the assumption on $|\Delta^t|$ we have

$$
\begin{aligned}
\sum_{j \in B_-^t} e^{\frac{-\alpha(\lceil -y_j^t \rceil + 1)}{n}} &\leqslant e^{-\frac{\alpha}{n}} \sum_{j \in B_-^t} e^{\frac{\alpha y_j^t}{n}} \\
&\leqslant e^{-\frac{\alpha}{n}} \cdot \left((|B_-^t| - 1) \cdot e^{-\frac{\alpha}{n} \cdot 0} + 1 \cdot e^{\frac{\alpha}{n} \cdot \sum_{j \in B_-^t} y_i^t} \right) \\
&\leqslant (|B_-^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{-\alpha/20},
\end{aligned}
$$

where we used the fact that $\sum_{j \in B_-^t} y_i^t = -\frac{1}{2}\Delta^t$. Applying this and the bound $e^x \leqslant 1+x+0.6 \cdot x^2$, for $x \leqslant 1/2$, to (5.6) gives

$$
\begin{aligned}
A_i &\leqslant (|B_-^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{-\alpha/20} + (|B_+^t| - 1) \cdot e^{-\frac{\alpha}{n}} + e^{\alpha} \\
&= (n-2) \cdot e^{-\frac{\alpha}{n}} + e^{-\alpha/20} + e^{\alpha} \\
&\leqslant (n-2) \cdot \left(1 - \frac{\alpha}{n} + \frac{6\alpha^2}{10n^2}\right) + \left(1 - \frac{\alpha}{20} + \frac{6\alpha^2}{4000}\right) + \left(1 + \alpha + \frac{6\alpha^2}{10}\right) \\
&= n \left(1 - \frac{\alpha(200 - 2406\alpha)}{4000n} + \mathcal{O}\left(\frac{\alpha^2}{n^2}\right)\right). \qquad (5.9)
\end{aligned}
$$

Thus we see by (5.8) and (5.9) that if $\mathcal{G}^t$ holds and we take $\alpha < 1/100$ and n sufficiently large, then there exists some constant $c_2 > 0$ such that $A_i \leqslant n(1 - c_2\alpha/n)$. Thus Lemma 5.1 gives

$$
\mathbf{E}[\, \Phi^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t \,] \leqslant \frac{1}{n} \sum_{i=1}^n \Phi_i^t \cdot A_i + e^{3\alpha} \leqslant \frac{1}{n} \cdot n \left(1 - \frac{c_2\alpha}{n}\right) \cdot \sum_{i=1}^n \Phi_i^t + e^{3\alpha} \leqslant \left(1 - \frac{c_2\alpha}{n}\right) \Phi^t + e^{3\alpha},
$$

as claimed. $\square$

The next lemma shows that the event $\mathcal{G}^t$ given by (5.4) holds for sufficiently many rounds.

Lemma 5.3. *Consider any allocation process satisfying $\mathcal{P}_1$ and $\mathcal{W}_1$. For every integer $t_0 \geqslant 1$, there are at least $n/40$ rounds $t \in [t_0, t_0 + n]$ with (i) $\Delta^t \geqslant 1/10 \cdot n$ or (ii) $|B_-^t| \geqslant n/20$.*

Proof. We claim that if $\Delta^s \leqslant 1/10 \cdot n$ for some round s, then for each round $t \in [s+n/5, s+9n/40]$ we have $|B_{-}^t| \geqslant n/20$ (deterministically). The lemma follows from that, since then we either have $\Delta^t \geqslant 1/10 \cdot n$ for all $t \in [t_0, t_0 + n/40]$, or, thanks to the claim there is a $s \in [t_0, t_0 + n/40]$ such that for all $t \in [s + n/5, s + 9n/40]$ (this interval has length $n/40$) we have $|B_{-}^t| \geqslant n/20$.

To establish the claim, assume we are at any round s where $\Delta^s \leqslant 1/10 \cdot n$. Then at most $n/2$ bins i satisfy $|y_i^s| \geqslant 1/5$, and in turn at least $n/2$ bins satisfy $|y_i^s| < 1/5$; let us call this latter set of bins $\mathcal{B} := \{i \in [n] : |y_i^s| < 1/5\}$. In the rounds $[s, s + 9n/40]$, we can choose at most $9n/40$ bins in $\mathcal{B}$ that are overloaded (at the time when chosen), and then we place exactly one ball into them. Furthermore, in each round $t \in [s, s + 9n/40]$ we can turn at most two bins in $\mathcal{B}$ which is underloaded at round t and make it overloaded. Hence it follows that at least $n/2 - 2 \cdot 9n/40 = n/20$ of the bins in $\mathcal{B}$ are not chosen in the interval $[s, s+9n/40]$. Consequently, these bins must be all underloaded in the interval $[s + n/5, s + 9n/40]$. $\square$

5.2 Completing the Proof of Theorem 4.3

We now introduce a new potential function $\widetilde{\Phi}^t$ which is the product of Φ^t with two additional terms (and an additive centering term). These multiplying terms have been chosen based on the one step increments in the two statements in Lemma 5.2 and Lemma 5.3. The purpose of this is that using Lemmas 5.2 and 5.3 we can show that $\widetilde{\Phi}^t$ is a super-martingale. We then use the super-martingale property to bound the exponential potential at an arbitrary step.

Here, we re-use the definition of the event $\mathcal{G}^t$ from (5.4). Now fix an arbitrary round $t_0 \geqslant 0$. Then, for any $s > t_0$, let $G_{t_0}^s$ be the number of rounds $r \in [t_0, s)$ satisfying $\mathcal{G}^r$, and let $B_{t_0}^s := (s - t_0) - G_{t_0}^s$. Further, let the constants $c_1 > 0$ and $c_2 > 0$ be as in Lemma 5.2, let $c_3 := 2e^{3\alpha} \exp(c_2\alpha) > 0$, and then define a sequence by $\widetilde{\Phi}^{t_0} := \Phi^{t_0}$, and for any $s > t_0$,

$$\widetilde{\Phi}^s := \Phi^s \cdot \exp\left(+\frac{c_2\alpha}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{s-1}\right) - c_3 \cdot (s - t_0). \tag{5.10}$$

The next lemma proves that the sequence $\widetilde{\Phi}^s$, $s \geqslant t_0$, forms a super-martingale:

Lemma 5.4. *Let $0 < \alpha < 1/100$ be an arbitrary but fixed constant, and $t_0 \geqslant 0$ be an arbitrary integer. Then, for any $s \in [t_0, t_0 + n]$ we have*

$$\mathbf{E}[\, \widetilde{\Phi}^{s+1} \mid \mathfrak{F}^s \,] \leqslant \widetilde{\Phi}^s.$$

Proof. First, recalling the definition $\widetilde{\Phi}^s$ from (5.10), we rewrite $\mathbf{E}[\, \widetilde{\Phi}^{s+1} \mid \mathfrak{F}^s \,]$ to give

$$\begin{aligned}
&\mathbf{E}[\, \widetilde{\Phi}^{s+1} \mid \mathfrak{F}^s \,] \\
&= \mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s \,] \cdot \exp\left(\frac{c_2\alpha}{n} \cdot G_{t_0}^s\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^s\right) - c_3 \cdot (s + 1 - t_0) \\
&= \mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s \,] \cdot \exp\left(\frac{\alpha}{n} \cdot (c_2 \cdot \mathbf{1}_{\mathcal{G}^s} - c_1\alpha \cdot (1 - \mathbf{1}_{\mathcal{G}^s}))\right) \cdot \exp\left(\frac{c_2\alpha}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{s-1}\right) \\
&\qquad - c_3 - c_3 \cdot (s - t_0).
\end{aligned}$$

We claim that it suffices to prove

$$\mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s \,] \cdot \exp\left(\frac{\alpha}{n} \cdot (c_2 \cdot \mathbf{1}_{\mathcal{G}^s} - c_1\alpha \cdot (1 - \mathbf{1}_{\mathcal{G}^s})\right) \leqslant \Phi^s + c_3 \cdot \exp\left(-c_2\alpha\right). \tag{5.11}$$

Indeed, observe that $G_{t_0}^{s-1} \leqslant s - t_0 \leqslant n$, and so assuming (5.11) holds we have

$$
\begin{aligned}
&\mathbf{E}[\, \widetilde{\Phi}^{s+1} \mid \mathfrak{F}^s \,] \\
&\leqslant (\Phi^s + c_3 \cdot \exp(-c_2\alpha)) \cdot \exp\left(\frac{c_2\alpha}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{s-1}\right) - c_3 - c_3 \cdot (s - t_0) \\
&= \Phi^s \cdot \exp\left(\frac{c_2\alpha}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{s-1}\right) + c_3 \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{s-1}\right) - c_3 - c_3 \cdot (s - t_0) \\
&\leqslant \widetilde{\Phi}^s.
\end{aligned}
$$

To show (5.11), we consider two cases based on whether $\mathcal{G}^s$ holds.
Case 1 [$\mathcal{G}^s$ holds]. By Lemma 5.2 (first statement) we have

$$
\mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s, \mathcal{G}^s \,] \leqslant \Phi^s \cdot \left(1 - \frac{c_2\alpha}{n}\right) + e^{3\alpha} \leqslant \Phi^s \cdot \exp\left(-\frac{c_2\alpha}{n}\right) + e^{3\alpha}.
$$

Hence, if $\mathcal{G}^s$ holds then the left hand side of (5.11) is equal to

$$
\begin{aligned}
\mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s, \mathcal{G}^s \,] \cdot \exp\left(\frac{\alpha}{n} \cdot c_2\right) &\leqslant \left(\Phi^s \cdot \exp\left(-\frac{c_2\alpha}{n}\right) + e^{3\alpha}\right) \cdot \exp\left(\frac{\alpha}{n} \cdot c_2\right) \\
&\leqslant \Phi^s + 2e^{3\alpha} \\
&= \Phi^s + c_3 \cdot \exp(-c_2\alpha),
\end{aligned}
$$

where the last line holds by definition of $c_3 = 2e^{3\alpha} \exp(c_2\alpha)$.
Case 2 [$\mathcal{G}^s$ does not hold]. Lemma 5.2 (second statement) gives the unconditional bound

$$
\mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s, \neg\mathcal{G}^s \,] \leqslant \Phi^s \cdot \left(1 + \frac{c_1\alpha^2}{n}\right) + e^{3\alpha} \leqslant \Phi^s \cdot \exp\left(\frac{c_1\alpha^2}{n}\right) + e^{3\alpha}.
$$

Thus, if $\mathcal{G}^s$ does not hold the left hand side of (5.11) is at most

$$
\mathbf{E}[\, \Phi^{s+1} \mid \mathfrak{F}^s, \neg\mathcal{G}^s \,] \cdot \exp\left(\frac{\alpha}{n} \cdot (-c_1\alpha)\right) \leqslant \Phi^s + e^{3\alpha} \leqslant \Phi^s + c_3 \cdot \exp(-c_2\alpha),
$$

which establishes (5.11) and the proof is complete. □

Combining Lemma 5.3, which shows that a constant fraction of rounds satisfy $\mathcal{G}^t$, with Lemma 5.4 establishes a multiplicative drop of $\mathbf{E}[\, \Phi^s \,]$ (unless it is already linear). Thus $\mathbf{E}[\, \Phi^m \,] = \mathcal{O}(n)$, which implies $\text{Gap}(m) = \mathcal{O}(\log n)$. This is formalized in the proof below.

Theorem 4.3 (restated). *Consider any allocation process which satisfies the conditions $\mathcal{P}_1$ and $\mathcal{W}_1$ at each round. Then, there is a constant $C > 0$ such that for any $m \geqslant 1$,*

$$
\mathbf{Pr}[\, \text{Gap}(m) \leqslant C \log n \,] \geqslant 1 - n^{-2}.
$$

Proof. For any integer $t_0 \geqslant 1$, first consider rounds $[t_0, t_0 + n]$. We will now fix the constant $\alpha := \min\{1/101, 1/20 \cdot c_2/c_1\}$ in the exponential potential function Φ^t (thus this is also fixed in $\widetilde{\Phi}^t$). By Lemma 5.4, $\widetilde{\Phi}$ forms a super-martingale over $[t_0, t_0 + n]$, and thus

$$
\mathbf{E}\left[\, \widetilde{\Phi}^{t_0+n} \,\middle|\, \mathfrak{F}^{t_0} \,\right] \leqslant \widetilde{\Phi}^{t_0} = \Phi^{t_0},
$$

which implies

$$
\mathbf{E}\left[\, \Phi^{t_0+n} \cdot \exp\left(+\frac{c_2\alpha}{n} \cdot G_{t_0}^{t_0+n-1}\right) \cdot \exp\left(-\frac{c_1\alpha^2}{n} \cdot B_{t_0}^{t_0+n-1}\right) - c_3 \cdot n \,\middle|\, \mathfrak{F}^{t_0} \,\right] \leqslant \Phi^{t_0},
$$

Rearranging this, and using that by Lemma 5.3, $G_{t_0}^{t_0+n-1} \geqslant n/40$ holds deterministically, we obtain for any $t_0 \geqslant 1$

$$\mathbf{E}\left[\, \Phi^{t_0+n} \mid \mathfrak{F}^{t_0} \,\right] \leqslant \left(\Phi^{t_0} + c_3 \cdot n\right) \cdot \exp\left(-c_2\alpha \cdot \frac{1}{40} + c_1\alpha^2 \cdot \frac{39}{40}\right),$$

and now using $\alpha = \min\{1/101, (1/40) \cdot c_2/c_1\}$ and defining $c_4 := c_2/40^2$ yields

$$\begin{aligned}\mathbf{E}\left[\, \Phi^{t_0+n} \mid \mathfrak{F}^{t_0} \,\right] &\leqslant \left(\Phi^{t_0} + c_3 \cdot n\right) \cdot \exp\left(-c_4\alpha\right) \\ &= \Phi^{t_0} \cdot \exp\left(-c_4\alpha\right) + c_3 \exp(-c_4\alpha) \cdot n.\end{aligned}$$

It now follows by the second statement in Lemma A.12 with $a := \exp(-c_4\alpha) < 1$ and $b := c_3 \exp(-c_4\alpha) \cdot n$ that for any integer $k \geqslant 1$,

$$\mathbf{E}\left[\, \Phi^{n \cdot k} \,\right] \leqslant \Phi^0 \cdot \exp\left(-c_4\alpha \cdot k\right) + \frac{c_3 \exp(-c_4\alpha) \cdot n}{1 - \exp(-c_4\alpha)} \leqslant c_5 \cdot n, \tag{5.12}$$

for some constant $c_5 > 0$ as $\Phi^0 \leqslant n$ holds deterministically.

Hence for any number of rounds $t = k \cdot n + r$, where $k \geqslant 0$ and $1 \leqslant r < n$, we use Lemma 5.2 (first statement) iteratively to conclude that

$$\begin{aligned}\mathbf{E}\left[\, \Phi^{n \cdot k + r} \,\right] &= \mathbf{E}\left[\, \mathbf{E}\left[\, \Phi^{n \cdot k + r} \mid \mathfrak{F}^{n \cdot k} \,\right] \,\right] \\ &\leqslant \mathbf{E}\left[\, \Phi^{n \cdot k} \,\right] \cdot \left(1 + \frac{c_1\alpha^2}{n}\right)^r + n \cdot \left(1 + \frac{c_1\alpha^2}{n}\right)^n \cdot e^{3\alpha} \\ &\leqslant c_5 \cdot n \cdot \exp(c_1\alpha^2) + n \cdot \exp(c_1\alpha^2) \cdot e^{3\alpha} \leqslant c_6 \cdot n,\end{aligned} \tag{5.13}$$

for some constant $c_6 > 0$. Hence for any $m \geqslant 1$, by Markov's inequality,

$$\mathbf{Pr}\left[\, \Phi^m \leqslant c_6 \cdot n^3 \,\right] \geqslant 1 - n^{-2}.$$

Since $\Phi^m \leqslant c_6 \cdot n^3$ implies $\text{Gap}(m) = \mathcal{O}(\log n)$, the proof is complete. □

5.3 Unfolding General Filling Processes

Recall the definition of the unfolding of a process from Page 11. First, we show that our notion of "unfolding" can be applied to capture the Caching process:

Lemma 4.4 (restated). *There is a process P satisfying conditions $\mathcal{P}_1$ and $\mathcal{W}_1$, such that for a suitable unfolding $U = U(P)$, the process U is an instance of* Caching.

Proof. As in the definition of unfolding, we denote the load vector of P after round t by x^t, and the load vector of a suitable unfolding $U(P)$ after the s-th atomic allocation by $\widehat{x}^s$. We also denote the corresponding normalized load vectors by y^t and $\widehat{y}^s$ respectively.

We will construct by induction, a coupling between a suitable filling process P, satisfying $\mathcal{W}_1$ and $\mathcal{P}_1$ at each round, and an unfolding $U(P)$ which follows the distribution of Caching. That is for every round $t \geqslant 0$ of P, there exists a (unique) atomic allocation $s = s(t) \geqslant t$ in $U(P)$, such that $x^t = \widehat{x}^{s(t)}$, and $U(P)$ is an instance of Caching.

Assume that for a suitable unfolding of the process P, the load configuration of P after round t equals the load configuration of $U(P)$ after atomic allocation $s = s(t)$, i.e., $x^t = \widehat{x}^{s(t)}$. In case the cache is empty (which happens only at the first round $s = 0$), Caching will sample a uniform bin i (which satisfies $\mathcal{P}_1$). If the cache is not empty, Caching will take as bin i the least loaded of the bin in the cache and a uniformly chosen bin. This produces a distribution vector that is majorized by one-choice (thus satisfies $\mathcal{P}_1$, again). Thus we may couple the two instances such that process P samples the same bin i in round t and atomic allocation $s(t)$, respectively. We continue with a case distinction concerning the load of bin i at round t:

Case 1 [The bin i is overloaded, i.e., $y_i^t = \widehat{y}_i^{s(t)} \geqslant 0$]. Then CACHING and P both place one ball into bin i, satisfying $\mathcal{W}_1$. Further, P proceeds to the next round and $U(P)$ proceeds to the next atomic allocation, which means that the coupling is extended.

Case 2 [The bin i is underloaded, i.e., $y_i^t = \widehat{y}_i^{s(t)} < 0$]. Then we can deduce by definition of CACHING that it will place the next $\lceil -y_i^{s(t)} \rceil + 1$ balls in some way that it deterministically satisfies the following conditions: (i) the first $\lceil -y_i^{s(t)} \rceil$ balls are placed into bins which have a normalized load < 0 at the atomic allocation $s(t)$, (ii) one ball is placed into a bin with normalized load < 1 at the atomic allocation $s(t)$. This follows since bin i gets stored in the cache and at each atomic allocation $j = W^{t-1} + 1, \ldots W^t$ the process has access to a cached bin with normalized load at most $y_i^{s(t)} + j - 1$. This satisfies $\mathcal{W}_1$ so we can continue the coupling.

We have thus constructed a process P, such that some unfolding $U = U(P)$ of P is an instance CACHING. $\square$

We now prove the general gap bound for unfolded processes.

Lemma 4.5 (restated). *Fix any constant $c > 0$. Then for any filling process P satisfying $\mathcal{W}_1$ and $\mathcal{P}_1$, there is a constant $C = C(c) > 0$ such that for any number of atomic allocations $m \geqslant 1$, with probability at least $1 - n^{-2}$, any unfolding $U = U(P)$ satisfies*

$$\left|\{t \in [m] \colon \operatorname{Gap}_U(t) \leqslant C \cdot \log n\}\right| \geqslant n^{-c} \cdot m \log m.$$

Proof. We will re-use the constants $\alpha \in (0,1)$ and $c_6 > 0$ defined in the proof of Theorem 4.3. We now define

$$\mathcal{B} := \left|\left\{t \in [1,m] \colon \Phi^t \geqslant c_6 \cdot n^{6+c}\right\}\right|,$$

which is the number of "bad" rounds of the filling process P. We continue with a case distinction for each round t whether $t \in \mathcal{B}$ holds.

- **Case 1** [$t \notin \mathcal{B}$]. By definition, for a round $t \notin B$ we have $\Phi^t < c_6 n^{6+c}$. Further, $\Phi^t < c_6 n^{6+c}$ implies $\operatorname{Gap}_P(t) < \frac{10}{\alpha} \cdot \log n$, for sufficiently large n. Now let $s(t), s(t) + 1, \ldots, s(t) + w^t - 1$, $w^t := \lceil -y_i^t \rceil + 1$ be the atomic allocations in $U(P)$ corresponding to round t in P. Since all allocations of $U(P)$ are to bins with normalized load at most 1 before the allocation, we conclude $\operatorname{Gap}_{U(P)}(s) \leqslant \max\{\operatorname{Gap}_P(t), 2\} < \frac{10}{\alpha} \cdot \log n$ for all $s \in [s(t), s(t) + w^t - 1]$.
- **Case 2** [$t \in \mathcal{B}$]. We will use that for any $0 < \alpha < 1$ and $t \geqslant 0$ we have
$$\Delta^t \leqslant 2n \cdot \left(\frac{1}{\alpha} \log \Phi^t + 1\right).$$
To see this, observe that $\sum_{i \in B_+^t} y_i^t = -\sum_{i \in B_-^t} y_i^t$ and thus
$$\Delta^t \leqslant 2 \sum_{i \in B_+^t} y_i^t \leqslant 2 \sum_{i \in B_+^t} y_i^t \mathbf{1}_{y_i^t \geqslant 2} + 2n. \tag{5.14}$$
Now, note that since $\Phi^t = \sum_{i \in [n] : y_i^t \geqslant 2} \exp\left(\alpha \cdot y_i^t\right)$, if $\Phi^t \leqslant \lambda$ then $y_i^t \leqslant \frac{1}{\alpha} \cdot \log \lambda$ for all $i \in [n]$ with $y_i^t \geqslant 2$. Thus by (5.14) we have $\Delta^t \leqslant 2n \cdot (1/\alpha) \log \Phi^t + 2n$ as claimed.
So for a round $t \in \mathcal{B}$, we have
$$w^t \leqslant \Delta^t + 1 \leqslant \frac{2n}{\alpha} \log \Phi^t + 2n + 1 \leqslant c_\alpha n \log \Phi^t, \tag{5.15}$$
which holds deterministically for some constant c_α. Again, every such round $t \in \mathcal{B}$ of P corresponds to the atomic allocations $s(t), s(t) + 1, \ldots, s(t) + w^t - 1$ in $U(P)$, and we will (pessimistically) assume that the gap in all those rounds is large, i.e., at least $\frac{10}{\alpha} \cdot \log n$.

By the above case distinction and (5.15), we can upper bound the number of rounds s in $[1, m]$ of $U(P)$ where $\mathrm{Gap}_{U(P)}(s) < \frac{10}{\alpha} \cdot \log n$ does not hold as follows:

$$\left| \left\{ s \in [1,m] \colon\ \mathrm{Gap}_{U(P)}(s) \geqslant \frac{10}{\alpha} \cdot \log n \right\} \right| \leqslant \sum_{t=1}^{m} \mathbf{1}_{t \in \mathcal{B}} \cdot w^t \leqslant c_\alpha n \cdot \sum_{t=1}^{m} \mathbf{1}_{t \in \mathcal{B}} \cdot \log \Phi^t. \tag{5.16}$$

Next define the sum of the exponential potential function over rounds 1 to m as

$$\Phi := \sum_{t=1}^{m} \Phi^t.$$

Then from the proof of Theorem 4.3, equation (5.13), there is a constant $c_6 > 0$ such that $\mathbf{E}\left[\, \Phi^t \,\right] \leqslant c_6 \cdot n$, and hence

$$\mathbf{E}\left[\, \Phi \,\right] = \sum_{t=1}^{m} \mathbf{E}\left[\, \Phi^t \,\right] \leqslant m \cdot c_6 \cdot n.$$

By Markov's inequality,

$$\mathbf{Pr}\left[\, \Phi \leqslant m \cdot c_6 \cdot n^3 \,\right] \geqslant 1 - n^{-2}.$$

Note that conditional on the above event occurring, the following bound holds deterministically:

$$|\mathcal{B}| \leqslant \frac{m c_6 n^3}{c_6 n^{6+c}} \leqslant m \cdot n^{-3-c}.$$

If $\mathcal{B}$ is empty we are done by (5.16). So we can assume that $|\mathcal{B}| \geqslant 1$ and apply the Log-sum inequality Lemma A.1, with $a_t = 1$ and $b_t = \Phi^t$, to the sum over $\mathcal{B}$ in (5.16), which gives us

$$\sum_{t=1}^{m} \mathbf{1}_{t \in \mathcal{B}} \cdot \log \Phi^t = \sum_{t \in \mathcal{B}} \log \Phi^t \leqslant \left(\sum_{t \in \mathcal{B}} 1 \right) \log \left(\frac{\sum_{t \in \mathcal{B}} \Phi^t}{\sum_{t \in \mathcal{B}} 1} \right) \leqslant |\mathcal{B}| \log \left(\frac{\Phi}{|\mathcal{B}|} \right). \tag{5.17}$$

Now, if the event $\Phi \leqslant m \cdot c_6 \cdot n^3$ occurs, then by (5.16) and (5.17) we have

$$\begin{aligned} \left| \left\{ s \in [1,m] \colon\ \mathrm{Gap}_{U(P)}(s) \geqslant \frac{10}{\alpha} \cdot \log n \right\} \right| &\leqslant c_\alpha n \cdot |\mathcal{B}| \cdot \log \left(\frac{\Phi}{|\mathcal{B}|} \right) \\ &\leqslant c_\alpha n \cdot (m \cdot n^{-3-c}) \cdot \log \left(m \cdot c_6 \cdot n^3 \right) \\ &\leqslant c'_\alpha m \cdot n^{-2-c} \cdot (\log m + \log n) \\ &\leqslant m \cdot n^{-c} \cdot \log m, \end{aligned}$$

where the third inequality is for some constant c'_α depending on α and the last inequality holds since $\alpha > 0$ is a small but fixed constant thus c'_α is constant. Since the inequality holds for *any* unfolding of P, once the event in (5.3) occurs, we obtain the corollary. □

5.4 Improved Analysis of the Caching Process

In this section, we will prove for the Caching process that w.h.p.[2] $\mathrm{Gap}(m) = \mathcal{O}(\log n)$, where $m \geqslant 1$ is arbitrary. This improves the guarantee we obtained in Section 5.3, which was based on regarding Caching as an "unfolded" version of a filling process.

We define an exponential potential Ψ^t which is similar to Φ^t, but with $\widetilde{\alpha} := \frac{1}{12n}$, and is given by

$$\Psi^t := \sum_{i : y_i^t \geqslant 2} \exp\left(\widetilde{\alpha} \cdot y_i^t \right) = \sum_{i=1}^{n} \exp\left(\widetilde{\alpha} \cdot y_i^t \right) \cdot \mathbf{1}_{\{y_i^t \geqslant 2\}}. \tag{5.18}$$

[2] In general, with high probability refers to probability of at least $1 - n^{-c}$ for some constant $c > 0$.

Our first goal will be to prove that $\mathbf{E}[\,\Psi^{t_0}\,] \leqslant 144n^3$ for an arbitrary round t_0, which in turn implies, using Markov's inequality, that w.h.p. $\Delta^{t_0} = \mathcal{O}(n^2 \log n)$. Note that this is weaker than the $\mathcal{O}(\log n)$ gap we are aiming to prove.

We will then use this gap as a starting point at $t_0 := m - n^7$ for a filling process satisfying $\mathcal{P}_1$ and $\mathcal{W}_1$, whose unfolding is Caching. Such a process exists by Lemma 4.4 and we will call it Folded-Caching in the rest of the section. For the potential function Φ defined in Section 5.1, we will show using, Theorem 4.3 and Lemma 4.5, that $\mathbf{E}[\,\Phi^{t_0+s}\,] = \mathcal{O}(n)$ for any $s \geqslant t_0 + n^3 \log^2 n$ rounds (which may unfold to multiple atomic allocations as the rounds of the Caching process). Finally, by taking Markov's inequality and the union bound over the next n^7 rounds which include the allocation of the m-th ball, we conclude that the exponential potential is small w.h.p. and hence $\mathrm{Gap}(m) = \mathcal{O}(\log n)$.

The most involved part of this proof is to prove that if Ψ^t is large, then it decreases in expectation. To this end, we will analyze two consecutive rounds of the Caching process. In these two rounds we show that even for the worst-case choice of b at the beginning of the two rounds, the Caching process has at least a $1/n^2$ bias to place the two balls away from the heaviest bin. This bias is enough to obtain a $\mathrm{poly}(n)$ gap. Note that over one step, the potential might increase in expectation even when large, e.g., when b stores the heaviest bin, in which case Caching allocates the next ball using One-Choice.

Lemma 5.5. *For any round $t \geqslant 0$, the* Caching *process satisfies*

$$\mathbf{E}[\,\Psi^{t+2} \mid \mathfrak{F}^t\,] \leqslant \Psi^t \cdot \Big(1 - \frac{1}{24n^3}\Big) + 6.$$

Proof. Recall that the filtration $\mathfrak{F}^t$ reveals the load vector x^t. Throughout this proof, we consider any labeling of the n bins such that $x_1^t \geqslant x_2^t \geqslant \cdots \geqslant x_n^t$ for round t. Now, consider a modified version of Caching, which we call Weak-Caching, which "forgets" the cache at time t and makes its decision for rounds t and $t+1$ using the loads x^t:

- In round t, place the ball in a bin i_1 sampled using One-Choice.
- In round $t+1$, sample another bin i_2 using One-Choice and place the ball in the least loaded of the i_1 and i_2, using the load information of round t. We break ties using the bin indices at round t.

Let $\widetilde{\Psi}$ be the potential (5.18) for Weak-Caching. The fact that Weak-Caching resets the cache implies that $\mathbf{E}[\,\Psi^{t+2} \mid x^t\,] \leqslant \mathbf{E}[\,\widetilde{\Psi}^{t+2} \mid x^t\,]+3$ (as we will show in Claim 5.6), while the fact that Weak-Caching uses outdated information from x^t allows us to analyze $\mathbf{E}[\,\widetilde{\Psi}^{t+2} \mid x^t\,]$.

Claim 5.6. *For any x^t and any choice of the cache b at round t,*

$$\mathbf{E}\left[\,\Psi^{t+2} \mid x^t\,\right] \leqslant \mathbf{E}\left[\,\widetilde{\Psi}^{t+2} \mid x^t\,\right] + 3.$$

Proof of Claim. Fix an arbitrary load vector x^t and consider the allocation of Weak-Caching and Caching in rounds t and $t+1$. We consider a coupling, where i_1 and i_2 are the bin samples of the two processes at rounds t and $t+1$. Let i_1^A, i_2^A and i_1^B, i_2^B be the pairs of bin indices where the Caching and Weak-Caching processes respectively allocate the two balls corresponding to t^{th} and $(t+1)^{\text{th}}$ allocation, and let x_A^{t+1}, x_A^{t+2} and x_B^{t+1}, x_B^{t+2} be the resulting load distributions. We will prove that x_B^{t+2} majorizes x_A^{t+2} by showing that $\langle x^{t+1}_{A,i_1^A}, x^{t+2}_{A,i_2^A}\rangle \leqslant \langle x^{t+1}_{B,i_1^B}, x^{t+2}_{B,i_2^B}\rangle$, where $\langle a, b\rangle := (\max\{a,b\}, \min\{a,b\})$. The claim will follow by Lemma A.6, since $2e^{3\alpha} < 3$.

The claim holds because Weak-Caching can allocate according to the best of two choices only for the second ball using outdated information, while Caching has this ability for the allocation of *both* balls. More formally, Weak-Caching will allocate to loads that are at least as large (because of the decisions being made on outdated information) as the two smallest loads in the multiset $M_B := \{x_{i_1}, x_{i_1}+1, x_{i_2}\}$. We consider the following cases (where each process allocates two balls) based on the cache b of Caching and the two choices i_1 and i_2.

- **Case A** [b, i_1, i_2 all different]. Caching allocates to the two smallest loads from the multiset $M_A = \{x_b, x_b + 1, x_{i_1}, x_{i_1} + 1, x_{i_2}\}$, while $M_B = \{x_{i_1}, x_{i_1} + 1, x_{i_2}\}$. So $M_B \subseteq M_A$.
- **Case B** [$b \neq i_1 = i_2$]. Caching allocates to the two smallest loads from the multiset $M_A = \{x_b, x_b + 1, x_{i_1}, x_{i_1} + 1\}$, while $M_B = \{x_{i_1}, x_{i_1} + 1\}$. So $M_B \subseteq M_A$.
- **Case C** [$b = i_1 \neq i_2$]. Caching allocates to the two smallest loads from the multiset $M_A = \{x_{i_1}, x_{i_1} + 1, x_{i_2}\}$, while $M_B = \{x_{i_1}, x_{i_1} + 1, x_{i_2}\}$. So $M_B \subseteq M_A$.
- **Case D** [$b = i_2 \neq i_1$]. Caching allocates to the two smallest loads from the multiset $M_A = \{x_{i_1}, x_{i_1} + 1, x_{i_2}, x_{i_2} + 1\}$, while $M_B = \{x_{i_1}, x_{i_1} + 1, x_{i_2}\}$. So $M_B \subseteq M_A$.
- **Case E** [$b = i_2 = i_1$]. Caching and Weak-Caching will allocate both balls to i_1.

In all these cases, $M_B \subseteq M_A$, so $\langle x^{t+1}_{A,i_1^A}, x^{t+2}_{A,i_2^A} \rangle \leqslant \langle x^{t+1}_{B,i_1^B}, x^{t+2}_{B,i_2^B} \rangle$, establishing the claim. $\Diamond$

Next we define $\widetilde{p}^t_{1,i}$ as the probability of allocating exactly one ball and $\widetilde{p}^t_{2,i}$ the probability of allocating two balls in bin i over the next two rounds t and $t+1$ using Weak-Caching. For any bin $i \in [n]$ with $y_i^t \geqslant 4 + 2/n$, we know that $y_i^{t+2} \geqslant 2$ so,

$$\begin{aligned}
\mathbf{E}[\, \widetilde{\Psi}_i^{t+2} \mid x^t \,] &= \widetilde{\Psi}_i^t \cdot e^{-2\widetilde{\alpha}/n} \cdot (1 - \widetilde{p}^t_{1,i} - \widetilde{p}^t_{2,i}) + \widetilde{\Psi}_i^t \cdot e^{\widetilde{\alpha} - 2\widetilde{\alpha}/n} \cdot \widetilde{p}^t_{1,i} + \widetilde{\Psi}_i^t \cdot e^{2\widetilde{\alpha} - 2\widetilde{\alpha}/n} \cdot \widetilde{p}^t_{2,i} \\
&= \widetilde{\Psi}_i^t \cdot e^{-2\widetilde{\alpha}/n} \cdot (1 + \widetilde{p}^t_{1,i} \cdot (e^{\widetilde{\alpha}} - 1) + \widetilde{p}^t_{2,i} \cdot (e^{2\widetilde{\alpha}} - 1)) \\
&= \Psi_i^t \cdot e^{-2\widetilde{\alpha}/n} \cdot (1 + \widetilde{p}^t_{1,i} \cdot (e^{\widetilde{\alpha}} - 1) + \widetilde{p}^t_{2,i} \cdot (e^{2\widetilde{\alpha}} - 1)).
\end{aligned}$$

Since for any i with $y_i^{t+2} < 2$, we have $\widetilde{\Psi}_i^{t+2} = 0$, for any bin $i \in [n]$ with $y_i^t \in [2, 4 + 2/n]$

$$\mathbf{E}[\, \widetilde{\Psi}_i^{t+2} \mid x^t \,] \leqslant \Psi_i^t \cdot e^{-2\widetilde{\alpha}/n} \cdot (1 + \widetilde{p}^t_{1,i} \cdot (e^{\widetilde{\alpha}} - 1) + \widetilde{p}^t_{2,i} \cdot (e^{2\widetilde{\alpha}} - 1)).$$

In rounds $t, t+1$ at most two bins can be change from $y_i^t < 2$ to $y_i^{t+2} > 2$. In this case, we have $y_i^{t+2} < 4$. So, their total contribution will be at most $2e^{\widetilde{\alpha}(3+2/n)} < 2e^{4\widetilde{\alpha}}$, so

$$\mathbf{E}[\, \widetilde{\Psi}^{t+2} \mid x^t \,] \leqslant \sum_{i : y_i^t \geqslant 2} \Psi_i^t \cdot e^{-2\widetilde{\alpha}/n} \cdot (1 + \widetilde{p}^t_{1,i} \cdot (e^{\widetilde{\alpha}} - 1) + \widetilde{p}^t_{2,i} \cdot (e^{2\widetilde{\alpha}} - 1)) + 2e^{4\widetilde{\alpha}}.$$

Using the Taylor estimate $e^z \leqslant 1 + z + z^2$ for $z < 1.75$ and that $2e^{4\widetilde{\alpha}} < 3$,

$$\mathbf{E}[\, \widetilde{\Psi}^{t+2} \mid x^t \,] \leqslant \sum_{i : y_i^t \geqslant 2} \Psi_i^t \cdot \left(1 - \frac{2\widetilde{\alpha}}{n} + \frac{4\widetilde{\alpha}^2}{n^2}\right) \cdot (1 + \widetilde{p}^t_{1,i} \cdot (\widetilde{\alpha} + \widetilde{\alpha}^2) + \widetilde{p}^t_{2,i} \cdot (2\widetilde{\alpha} + 4\widetilde{\alpha}^2)) + 3.$$

For Weak-Caching, the probability of allocating exactly one ball to bin $i \in [n]$ (with respect to the load ordering at round t) over the next two rounds is,

$$\widetilde{p}^t_{1,i} = \mathbf{Pr}\left[\, \{i_1 = i\} \cap \{i_2 > i\} \,\right] + \mathbf{Pr}\left[\, \{i_1 < i\} \cap \{i_2 = i\} \,\right] = \frac{1}{n} \cdot \left(1 - \frac{i}{n}\right) + \frac{i-1}{n} \cdot \frac{1}{n} = \frac{1}{n} - \frac{1}{n^2},$$

where we have used the fact that Weak-Caching allocates the two balls using the information of the load vector x^t only, and break ties according to the ball index of the (sorted) load vector x^t. Similarly, the probability of allocating two balls is,

$$\widetilde{p}^t_{2,i} = \mathbf{Pr}\left[\, \{i_1 = i\} \cap \{i_2 \leqslant i\} \,\right] = \frac{1}{n} \cdot \frac{i}{n}.$$

Let $\overline{p}^t_2 := \frac{1}{|\{i : y_i^t \geqslant 2\}|} \cdot \sum_{i : y_i^t \geqslant 2} \widetilde{p}^t_{2,i}$ be the average probability of allocating two balls to bin i with $y_i^t \geqslant 2$. Note that since $\widetilde{p}^t_{2,i}$ is increasing in i, the larger $|\{i : y_i^t \geqslant 2\}|$ is, the larger $\overline{p}^t_2$ will be. Since there can be at most $n-1$ bins in with $y_i^t \geqslant 2$, we have that $|\{i : y_i^t \geqslant 2\}| \leqslant n - 1$, hence,

$$\overline{p}^t_2 \leqslant \frac{1}{n^2} \cdot \frac{1}{n-1} \cdot \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2n^2(n-1)} = \frac{1}{2n}.$$

Now, the expected change for the Ψ potential over the next two rounds is at most,

$$\begin{aligned}
\mathbf{E}[\,\Psi^{t+2} \mid \mathcal{F}^t\,] &= \mathbf{E}[\,\Psi^{t+2} \mid \mathcal{F}^t, x^t\,] \\
&\overset{(a)}{\leqslant} \mathbf{E}[\,\widetilde{\Psi}^{t+2} \mid x^t\,] + 3 \\
&\leqslant \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{2\widetilde{\alpha}}{n} + \frac{4\widetilde{\alpha}^2}{n^2}\Big) \cdot (1 + \widetilde{p}_{1,i}^t \cdot (\widetilde{\alpha} + \widetilde{\alpha}^2) + \widetilde{p}_{2,i}^t \cdot (2\widetilde{\alpha} + 4\widetilde{\alpha}^2)) + 6 \\
&\overset{(b)}{\leqslant} \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{2\widetilde{\alpha}}{n} + \frac{4\widetilde{\alpha}^2}{n^2}\Big) \cdot (1 + \widetilde{p}_{1,i}^t \cdot (\widetilde{\alpha} + \widetilde{\alpha}^2) + \overline{p}_{2}^t \cdot (2\widetilde{\alpha} + 4\widetilde{\alpha}^2)) + 6 \\
&\leqslant \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{2\widetilde{\alpha}}{n} + \frac{4\widetilde{\alpha}^2}{n^2}\Big) \cdot \Big(1 + \Big(\frac{1}{n} - \frac{1}{n^2}\Big) \cdot (\widetilde{\alpha} + \widetilde{\alpha}^2) + \frac{1}{2n} \cdot (2\widetilde{\alpha} + 4\widetilde{\alpha}^2)\Big) + 6 \\
&= \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{2\widetilde{\alpha}}{n} + \frac{4\widetilde{\alpha}^2}{n^2}\Big) \cdot \Big(1 + \frac{2\widetilde{\alpha}}{n} - \frac{\widetilde{\alpha}}{n^2} + \frac{3\widetilde{\alpha}^2}{n} - \frac{\widetilde{\alpha}^2}{n^2}\Big) + 6 \\
&\overset{(c)}{=} \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{2\widetilde{\alpha}}{n} + \frac{2\widetilde{\alpha}}{n} - \frac{\widetilde{\alpha}}{n^2} + \frac{3\widetilde{\alpha}^2}{n} + o(n^{-3})\Big) + 6 \\
&= \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{\widetilde{\alpha}}{n^2} + \frac{3\widetilde{\alpha}^2}{n} + o(n^{-3})\Big) + 6 \\
&\overset{(d)}{\leqslant} \sum_{i:y_i^t \geqslant 2} \Psi_i^t \cdot \Big(1 - \frac{1}{24n^3}\Big) + 6,
\end{aligned}$$

where (a) uses Claim 5.6, (b) uses majorization (Lemma A.2), since Ψ_i^t is non-decreasing and $\widetilde{p}_{2,i}^t$ is increasing and (c) as well as (d) use that $\widetilde{\alpha} = 1/(12n)$. ☐

Lemma 5.7. *For any round $t \geqslant 0$ of the* Caching *process we have*

$$\mathbf{E}[\,\Psi^t\,] \leqslant 144n^3.$$

Proof. We will prove the statement by strong induction. For $t = 0$ and $t = 1$, we have $\Psi^0 = 0$ and $\Psi^1 = 0$ deterministically, so the statement holds. Assume $\mathbf{E}[\,\Psi^s\,] \leqslant 144n^3$ for all $s \leqslant t+1$, then using Lemma 5.5,

$$\mathbf{E}[\,\Psi^{t+2}\,] = \mathbf{E}[\,\mathbf{E}[\,\Psi^{t+2} \mid \mathfrak{F}^t\,]\,] \leqslant \mathbf{E}[\,\Psi^t\,] \cdot \Big(1 - \frac{1}{24n^3}\Big) + 6 \leqslant 144n^3 - \frac{144n^3}{24n^3} + 6 = 144n^3. \quad ☐$$

Lemma 5.8. *For the* Caching *process, for any round $t \geqslant 0$,*

$$\mathbf{Pr}\Big[\max_{i \in [n]} y_i^t \leqslant 14 \cdot n \log n\Big] \geqslant 1 - n^{-9}.$$

Proof. Using Lemma 5.7 and Markov's inequality, we get,

$$\mathbf{Pr}\big[\,\Psi^t \leqslant 144n^{13}\,\big] \geqslant 1 - n^{-9}.$$

Since $\Psi^t \leqslant 144n^{13}$ implies $\min_{i \in [n]} y_i^t \leqslant 13 \cdot n \log n + \log 144 \leqslant 14 \cdot n \log n$, we conclude that

$$\mathbf{Pr}\Big[\max_{i \in [n]} y_i^t \leqslant 14 \cdot n \log n\Big] \geqslant 1 - n^{-9}. \quad ☐$$

The next claim holds not only for Folded-Caching, but more generally for any allocation process satisfying $\mathcal{W}_1$, and we will make use of the more general claim in Section 10.

Claim 5.9. *For any allocation process satisfying $\mathcal{W}_1$, it holds for any $t \geqslant 0$ that $\Delta^{t+1} \leqslant \Delta^t + 4$.*

Proof. We have to relate Δ^t to Δ^{t+1} where we recall that, for any $t \geqslant 0$,

$$\Delta^t := \sum_{i \in [n]} \left| x_i^t - \frac{W^t}{n} \right|.$$

To do this, let $i \in [n]$ be the chosen bin for the allocation in round t.

Case 1 $[y_i^t \geqslant 0]$. If $i \in [n]$ satisfies $y_i^t \geqslant 0$, i.e., the chosen bin is overloaded, then we place exactly one ball into i. This increases the average load by $\frac{1}{n}$ and increments the load of exactly one bin by 1. By using the triangle inequality this implies

$$\Delta^{t+1} \leqslant \Delta^t + 2.$$

Case 2 $[y_i^t < 0]$. If bin i is underloaded, then we allocate exactly $w^t := \lceil -y_i^t \rceil + 1$ many balls in round t. We now split the allocation of these w^t balls within round t into w^t atomic allocations labeled $1, 2, \ldots, w^t$. Correspondingly, we define the absolute value potential for atomic allocations between t and $t+1$, by setting for any $\ell \in [0, w^t]$

$$\Delta^{t,\ell} := \sum_{i \in [n]} \left| x_i^{t,\ell} - \frac{W^t + \ell}{n} \right|,$$

where $x_{t,\ell}$ is the load vector after ℓ-th atomic allocation of round t. Since the order of the placement of the balls does not matter for $\Delta^{t,w^t} = \Delta^{t+1}$, we may assume that the bins k_1 and k_2 in the definition of $\mathcal{W}_1$ (if they exist), are used in the first and second atomic allocation of round t, respectively. Using the same as argument as in Case 1, we conclude that $\Delta^{t,1} \leqslant \Delta^{t,0}+2$, and $\Delta^{t,2} \leqslant \Delta^{t,1} + 2$, so that

$$\Delta^{t,2} \leqslant \Delta^{t,0} + 4.$$

Further, note that any bin $j \in [n]$ apart from k_1 and k_2 receives at most $\lceil -y_j^t \rceil - 1$ many balls in round t. Hence if we allocate to a bin j in the ℓ-th atomic allocation, where $\ell \geqslant 3$, it can have received at most $\big(\lceil -y_j^t \rceil - 1\big) - 1 = \lceil -x_j^t + \frac{W^t}{n} \rceil - 2$ many balls in previous atomic allocations. Hence

$$x_j^{t,\ell} - \frac{W^t + \ell}{n} \leqslant x_j^{t,\ell} - \frac{W^t}{n} \leqslant x_j^t + \left\lceil -x_j^t + \frac{W^t}{n} \right\rceil - 2 - \frac{W^t}{n} \leqslant -1.$$

Therefore,

$$\left| x_j^{t,\ell+1} - \frac{W^t + (\ell+1)}{n} \right| = \left| \underbrace{x_j^{t,\ell} - \frac{W^t + \ell}{n}}_{\leqslant -1} +1 - \frac{1}{n} \right| = \left| x_j^{t,\ell} - \frac{W^t + \ell}{n} \right| - \left(1 - \frac{1}{n}\right). \tag{5.19}$$

On the other hand, for any $j \neq i$, we have by the triangle inequality

$$\begin{aligned}\sum_{j \in [n], j \neq i} \left| x_j^{t,\ell+1} - \frac{W^t + (\ell+1)}{n} \right| &= \sum_{j \in [n], j \neq i} \left| x_j^{t,\ell} - \frac{W^t + \ell}{n} + \frac{1}{n} \right| \\ &\leqslant \sum_{j \in [n], j \neq i} \left| x_j^{t,\ell} - \frac{W^t + \ell}{n} \right| + (n-1) \cdot \frac{1}{n}. \end{aligned} \tag{5.20}$$

If we combine (5.19) and (5.20) then the $-(1 - 1/n)$ and $+(n-1)/n$ terms cancel, so we have

$$\Delta^{t,\ell+1} \leqslant \Delta^{t,\ell}.$$

Aggregating the changes over $\ell = 0, 1, 2, \ldots, w^t$, we conclude

$$\Delta^{t+1} = \Delta^{t,w^t} \leqslant \Delta^{t,0} + 4 = \Delta^t + 4. \qquad \square$$

We now (re-)state and prove the main result of this section: the gap bound for Caching.

Lemma 4.6 (restated). *For the* Caching *process (which allocates exactly one ball per round) there is a constant $C > 0$ such that for any number of rounds $m \geqslant 1$,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant C \cdot \log n\,\right] \geqslant 1 - n^{-2}.$$

Proof. First note that if $m < n^7$, then the statement of the lemma is also covered by Lemma 4.5 with $c = 8$. Thus we will assume $m \geqslant n^7$ in the following.

In order to analyze the state of Caching at round m, we will analyze Caching in two different phases. The first phase lasts from round 1 to $m - n^7$, and in this phase we consider the (original) Caching process. We use the above analysis to prove a (polynomial) bound on the potential Ψ for round $m - n^7$, which implies that $\Delta^t \leqslant 14n^2 \log n$ and an initial bound on Φ^{m-n^7}. The second phase starts with the load configuration at round $m - n^7$, and then considers n^7 rounds of Folded-Caching (as a process which satisfies $\mathcal{P}_1$ and $\mathcal{W}_1$). The folded process allocates at least one ball (and possibly more) at each round, and hence the allocation of the m-th ball comes before round m in the second phase. Also, as we will prove below, the allocation of the m-th ball does not happen before round $m - \frac{1}{2}n^{3.5}$, which gives enough time for the exponential potential Φ (used in the analysis in Section 5) to decrease.

Define $t_0 := m - n^7$, which is the last round of phase one when we switch from Caching to Folded-Caching. Following Lemma 5.8, we define an event

$$\mathcal{G}^{t_0} := \left\{ \max_{i \in [n]} y_i^{t_0} \leqslant 14 \cdot n \log n \right\}$$

and thus at at round t_0, we have

$$\mathbf{Pr}\left[\,\mathcal{G}^{t_0}\,\right] \geqslant 1 - n^{-9}. \tag{5.21}$$

If $\mathcal{G}^{t_0}$ occurs, then the following two bounds on potential functions hold:

$$\Phi^{t_0} \leqslant n \cdot \exp(\alpha \cdot 14 \cdot n \log n), \tag{5.22}$$

and

$$\Delta^{t_0} \leqslant 14 \cdot n^2 \log n. \tag{5.23}$$

Fix now any filtration $\mathcal{F}^{t_0}$, including the load vector x^{t_0}, such that $\mathcal{G}^{t_0}$ holds. We will now look at the process after round t_0, as mentioned before, we will consider Folded-Caching process (i.e., multiple balls may be allocated in each round). For convenience, the time-indices larger than t_0 will refer to the rounds of Folded-Caching the instead of atomic allocations. Thus, for any integer $s \geqslant 0$, x^{t_0+s} refers to the resulting load vector after s allocations of Folded-Caching have been completed, starting with x^{t_0}.

Next consider $s \geqslant n^3 \log^2 n > \log(n \cdot \exp(14 \cdot n \log n))$ rounds after t_0 and write $s = \kappa_1 \cdot n + \kappa_2$ for $\kappa_1, \kappa_2 \in \mathbb{N}$ and $0 \leqslant \kappa_2 < n$. Then,

$$\begin{aligned}
\mathbf{E}\left[\,\Phi^{t_0+n\cdot\kappa_1} \mid \mathfrak{F}^{t_0}\,\right], \mathcal{G}^{t_0} &\overset{(a)}{\leqslant} \Phi^{t_0} \cdot e^{-c_4\alpha\cdot\kappa_1} + \sum_{i=0}^{\kappa_1-1} c_3 e^{-c_4\alpha} \cdot e^{-c_4\alpha\cdot i} \\
&\overset{(b)}{\leqslant} n \cdot e^{14\cdot n\log n} \cdot e^{-c_4\alpha\cdot\lfloor n^2\log^2 n\rfloor} + \sum_{i=0}^{\kappa_1-1} c_3 e^{-c_4\alpha} \cdot e^{-c_4\alpha\cdot i} \leqslant c_5 n,
\end{aligned}$$

for some constant $c_5 > 0$, where (a) used (5.12) and (b) used (5.22).

Similarly to (5.13),

$$\begin{aligned}\mathbf{E}\left[\, \Phi^{t_0+\kappa_1\cdot n+\kappa_2} \;\middle|\; \mathfrak{F}^{t_0}, \mathcal{G}^{t_0} \,\right] &\leqslant \mathbf{E}\left[\, \Phi^{\kappa_1 \cdot n} \;\middle|\; \mathfrak{F}^{t_0}, \mathcal{G}^{t_0} \,\right] \cdot \left(1 + \frac{c_1\alpha^2}{n}\right)^{\kappa_2} + n \cdot \left(1 + \frac{c_1\alpha^2}{n}\right)^{n} \cdot e^{3\alpha} \\ &\leqslant c_5 \cdot n \cdot \exp(c_1\alpha^2) + n \cdot \exp(c_1\alpha^2) \cdot e^{3\alpha} \leqslant c_6 \cdot n,\end{aligned}$$

for some constant $c_6 > 0$. Note that in any round t, Folded-Caching allocates at most $\Delta^t + 4$ balls. Hence during the s rounds after t_0 of Folded-Caching, the total number of balls allocated is at most

$$\begin{aligned}(\Delta^{t_0} + 4) + (\Delta^{t_0+1} + 4) + \ldots + (\Delta^{t_0+s} + 4) &\leqslant (\Delta^{t_0} + 4) + (\Delta^{t_0} + 8) + \ldots + (\Delta^{t_0} + 4 \cdot (s+1)) \\ &\leqslant \Delta^{t_0} \cdot (s+1) + 2(s+1)(s+2),\end{aligned}$$

where we used that for any round $t \geqslant t_0$, $\Delta^{t+1} \leqslant \Delta^t + 4$ (Claim 5.9).

So, for any $s < \frac{1}{2}n^{3.5}$, the total number of balls allocated in the s rounds after t_0 is at most

$$\Delta^{t_0} \cdot (s+1) + 2(s+1)(s+2) < (14 \cdot n^2 \log n) \cdot \left(\frac{1}{2}n^{3.5} + 1\right) + \frac{1}{2}n^7 \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) < n^7,$$

where we used (5.23) in the first inequality. Since the first t_0 rounds allocate $m - n^7$ balls in total, this means that we do not allocate the m-th ball before the s-th round after t_0. Also since Folded-Caching allocates at each round at least one ball, we allocate the m-th ball at some round r in $\mathcal{R} := [\frac{1}{2}n^{3.5}, n^7]$ after t_0.

Now applying Markov's inequality we have, for any fixed $r \in \mathcal{R}$,

$$\mathbf{Pr}\left[\, \Phi^r \leqslant c_6 n^{11} \;\middle|\; \mathcal{G}^{t_0} \,\right] \geqslant 1 - n^{-10}.$$

By taking the union bound over the rounds $\mathcal{R}$ (with $|R| < n^7$),

$$\mathbf{Pr}\left[\, \bigcap_{s \in \mathcal{R}} \left\{\Phi^{t_0+s} \leqslant c_6 n^{11}\right\} \;\middle|\; \mathcal{G}^{t_0} \,\right] \geqslant 1 - n^7 \cdot n^{-10} \geqslant 1 - n^{-3}.$$

Note that $\Phi^{t_0+s} \leqslant c_6 n^{11}$ implies that for some constant $C > 0$, $x_{\max}^{t_0+s} \leqslant C \log n$. Using this,

$$\mathbf{Pr}\left[\, \bigcap_{s \in \mathcal{R}} \left\{x_{\max}^{t_0+s} \leqslant C \log n\right\} \;\middle|\; \mathcal{G}^{t_0} \,\right] \geqslant 1 - n^{-3}.$$

Finally, since if $\mathcal{G}^{t_0}$ holds, the round when we throw the m-th ball is in $\mathcal{R}$, by the relationship between the Folded-Caching and Caching in Lemma 4.4 for the last n^7 rounds, we conclude from the above and (5.21),

$$\begin{aligned}\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant C \log n \,\right] &= \mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant C \log n \;\middle|\; \mathcal{G}^{t_0} \,\right] \cdot \mathbf{Pr}\left[\, \mathcal{G}^{t_0} \,\right] \\ &\geqslant (1 - n^{-3}) \cdot (1 - n^{-9}) \geqslant 1 - n^{-2}.\end{aligned}$$

□

6 Overview of the Analysis of Non-Filling Processes

The analysis for non-filling processes re-uses some ideas from Section 5, but is substantially more involved (a diagram summarizing all key steps in the analysis is given in Fig. 5). The reason for this complexity comes from the inability to fill "big holes", i.e., even if a bin with extremely small load is sampled, only a constant number (e.g. two for Twinning) of balls will be allocated (due to $\mathcal{W}_3$). This is not only a challenge for the analysis, but it also leads to a completely different behavior of the process, in comparison to filling processes studied in Section 5.

All of the following analysis in this section (and Sections 7 to 9) will be for processes which satisfy $\mathcal{P}_2$ and $\mathcal{W}_2$. If the stronger condition (which is the precondition of Theorem 4.15) that additionally at least one of $\mathcal{P}_3$ or $\mathcal{W}_3$ hold, then we state this explicitly by referring to a $\mathcal{P}_3 \cap \mathcal{W}_2$-process, or, $\mathcal{P}_2 \cap \mathcal{W}_3$-process.

Our analysis will study the interplay between the following potential functions.

- The *absolute value potential* (this is also known as the *number of holes* in [2]):

$$\Delta^t := \sum_{i=1}^n |y_i^t|.$$

  In Lemma 6.1, we prove that when $\Delta^t = \mathcal{O}(n)$, then w.h.p., among next $\Theta(n)$ rounds a constant fraction of them are rounds t with $\delta^t \in (\varepsilon, 1-\varepsilon)$, for some constant $\varepsilon > 0$.

- The *quadratic potential*:

$$\Upsilon^t := \sum_{i=1}^n \left(y_i^t\right)^2.$$

  In Lemma 6.2, we prove that for processes satisfying $\mathcal{W}_3$ or $\mathcal{P}_3$, if for some round t, $\Delta^t = \Omega(n)$ holds, then Υ^t decreases in expectation.

- The exponential potential function for a constant $\alpha > 0$ to be specified later:

$$\Lambda^t := \sum_{i=1}^n \exp\left(\alpha \cdot |y_i^t|\right) = \sum_{i \in B_+^t} \exp\left(\alpha y_i^t\right) + \sum_{i \in B_-^t} \exp\left(-\alpha(-y_i^t)\right).$$

  In Corollary 8.6, we prove that for $\Lambda^t = \Omega(n)$, if $\delta^t \in (\varepsilon, 1-\varepsilon)$, then Λ^t decreases in expectation, otherwise it increases by a smaller factor (Corollary 8.6). Note that unlike the analysis of the $(1+\beta)$-process in [24] and unlike one batch of n balls in filling processes, the potential can increase in expectation over a single round, even when it is large (Claim B.2).

- A "weaker" instance of the Λ potential function where $\widetilde{\alpha} = \Theta(1/n)$:

$$V^t := \sum_{i=1}^n \exp\left(\widetilde{\alpha} \cdot |y_i^t|\right) = \sum_{i \in B_+^t} \exp\left(\widetilde{\alpha} y_i^t\right) + \sum_{i \in B_-^t} \exp\left(-\widetilde{\alpha}(-y_i^t)\right).$$

  For this potential, we can directly establish that $\mathbf{E}[\, V^t \,] = \operatorname{poly}(n)$ at an arbitrary round t, without considering the quantile δ^t. Then, using Markov's inequality we establish w.h.p. that $\operatorname{Gap}(t) = \mathcal{O}(n \log n)$. Note that this is similar to the case $\beta = \Theta(1/n)$ in the $(1+\beta)$ process [24]. We use this as the base case of the heavily-loaded case.

Having defined the potential functions, let us now describe in more detail why the exponential potential function Λ may increase in expectation in some of the rounds.

While for filling processes, from *any* load configuration, the exponential potential function goes down every $\mathcal{O}(n)$ rounds (unless it is already small), this is not true for, e.g., the Thinning process. There exist bad configurations where (*i*) $\operatorname{Gap}(t)$ is large and (*ii*) for all $s \in [t, t+\omega(n)]$ rounds $\delta^s = 1 - o(1)$ (or $\delta^s = o(1)$), where δ^s is the quantile of the mean load at round t. Note that if δ^s is too close to 1 (or 0), then the bias away from any *fixed* overloaded (or towards any *fixed* underloaded) bin is too small, and the process allocates balls almost uniformly, similarly to One-Choice. As a result, the exponential potential may increase for several rounds, until δ^s is bounded away from 0 and 1 (see Fig. 6 for an illustration showing experimental results and Claim B.2 for a concrete example).

So, instead we start with the weaker exponential potential function V where $\widetilde{\alpha} = \Theta(1/n)$. Because non-filling processes have a small bias to place away from the maximum load, we are able to prove in Section 8.3 that when V is sufficiently large it decreases in expectation. This allows us to prove that $\mathbf{E}[\,V^t\,] = \mathrm{poly}(n)$ at an arbitrary round t and infer, using Markov's inequality, that the gap is w.h.p. $\mathrm{Gap}(t) = \mathcal{O}(n \log n)$ (Lemma 8.10). Then, our next goal is to show that starting with $\mathrm{Gap}(t_0) = \mathcal{O}(n \log n)$ we reach $s \in [t_0, t_0 + \Theta(n^3 \log^4 n)]$ where $\Lambda^s = \mathcal{O}(n)$ (the *recovery phase* – Section 9.3) and finally show that the gap remains $\mathcal{O}(\log n)$ for steps $[s, m]$ (the *stabilization phase* – Section 9.4).

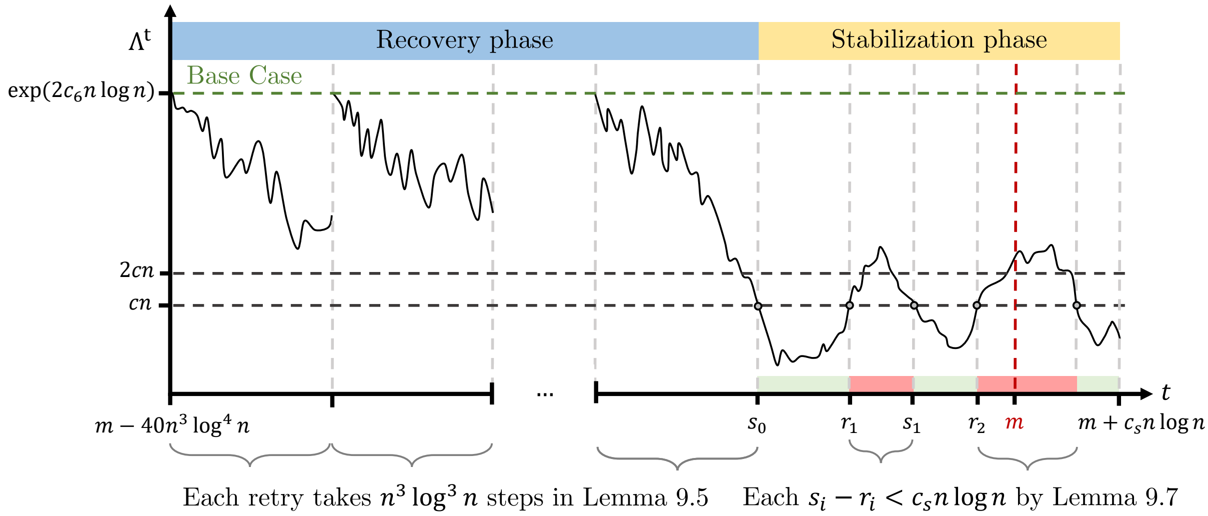

Figure 4: Overview of the recovery phase and stabilization phase. In the recovery phase starting with $V^t = \mathrm{poly}(n)$ (and so $\mathrm{Gap}(t) = \mathcal{O}(n \log n)$), we will perform several retries, each consisting of $n^3 \log^3 n$ consecutive steps until we find a round s_0 with $\Lambda^{s_0} < cn$. Each retry is successful with constant probability > 0, such after at most $\mathcal{O}(\log n)$ retries we will have w.h.p. one success. We then switch to the stabilization phase, where in Lemma 9.7 we prove that every $c_s n \log n$ rounds $\Lambda^s < cn$ is satisfied, which allows us to infer that $\mathrm{Gap}(m) = \mathcal{O}(\log n)$.

In both the recovery and the stabilization phase, we will study the evolution of δ^t and prove that from any load configuration, the process eventually reaches a value δ^t in $(\varepsilon, 1-\varepsilon)$ for some constant $\varepsilon > 0$, sufficiently often. The next lemma provides a sufficient condition for the process at round t_1 to reach a phase $[t_1, t_1 + \Theta(n)]$ where $\delta^t \in (\varepsilon, 1-\varepsilon)$ occurs often:

Lemma 6.1 (Mean Quantile Stabilization). *Consider any allocation process satisfying $\mathcal{P}_2$ and $\mathcal{W}_2$. Then for any integer constant $C \geqslant 1$, there exists some $\varepsilon := \varepsilon(C) > 0$ such that for any integers $t_0 \geqslant 0$ and $t_3 := t_0 + \left\lceil \frac{2Cn}{w_+} \right\rceil + \left\lceil \frac{n}{w_+} \right\rceil + \left\lceil \frac{n}{10 w_-} \right\rceil$ we have*

$$\mathbf{Pr}\left[\, \left| \left\{ t \in [t_0, t_3] \colon \delta^t \in (\varepsilon, 1-\varepsilon) \right\} \right| \geqslant \varepsilon \cdot n \;\middle|\; \mathfrak{F}^{t_0}, \Delta^{t_0} \leqslant C \cdot n \,\right] \geqslant 1 - e^{-\varepsilon \cdot n}.$$

Thus, whenever the absolute value potential Δ^{t_1} is at most linear, the quantile δ^t is 'good' (bounded away from 0 or 1) in a constant fraction of the next $\Theta(n)$ rounds. The next step in the proof is to establish that the sufficient condition on Δ^{t_1} will be satisfied. To this end, we use a relation between the absolute value potential Δ^t and the quadratic potential Υ^t, showing that Υ^t drops in expectation as long as $\Delta^t = \Omega(n)$. Thus, Δ^t must eventually become linear.

Lemma 6.2. *Consider any allocation process satisfying $\mathcal{W}_2$ and $\mathcal{P}_2$. Then for any $t \geqslant 0$, the quadratic potential satisfies*

$$\mathbf{E}\left[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - (p_-^t \cdot w_- - p_+^t \cdot w_+) \cdot \Delta^t + 4 \cdot (w_-)^2.$$

Hence for any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process, this implies by Lemma 4.14 that there exist constants $c_1, c_2 > 0$ such that for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - \frac{c_1}{n} \cdot \Delta^t + c_2.$$

Combining the above two lemmas, we prove that for a constant fraction rounds the mean quantile δ^t is good. In particular, for the recovery phase, we show this guarantee holds with constant probability for an interval of length $\Theta(n \log^3 n)$ (Lemma 9.3) and for the stabilization phase, we prove that it holds w.h.p. for an interval of length $\Theta(n \log n)$ given that we start with $\Lambda^s = \mathcal{O}(n)$ (Lemma 9.4). In these good rounds, similarly to Lemma 5.2, we prove that the exponential potential Λ^t decreases by $(1 - c_3\alpha/n)$ (see Lemma 8.4). In other rounds, the potential Λ^t increases by $(1 + c_5\alpha^2/n)$ (see Lemma 8.5). Combining these, we obtain that the exponential potential function eventually becomes $\mathcal{O}(n)$, which implies a logarithmic gap. For the recovery phase, we use a "retry" argument to amplify the probability from constant to $1 - n^{-\Omega(1)}$. We refer to Fig. 4 for a high-level overview of recovery and stabilization, as well as Fig. 5 for a diagram summarizing most of the crucial lemmas used in the analysis and outlining their relationship.

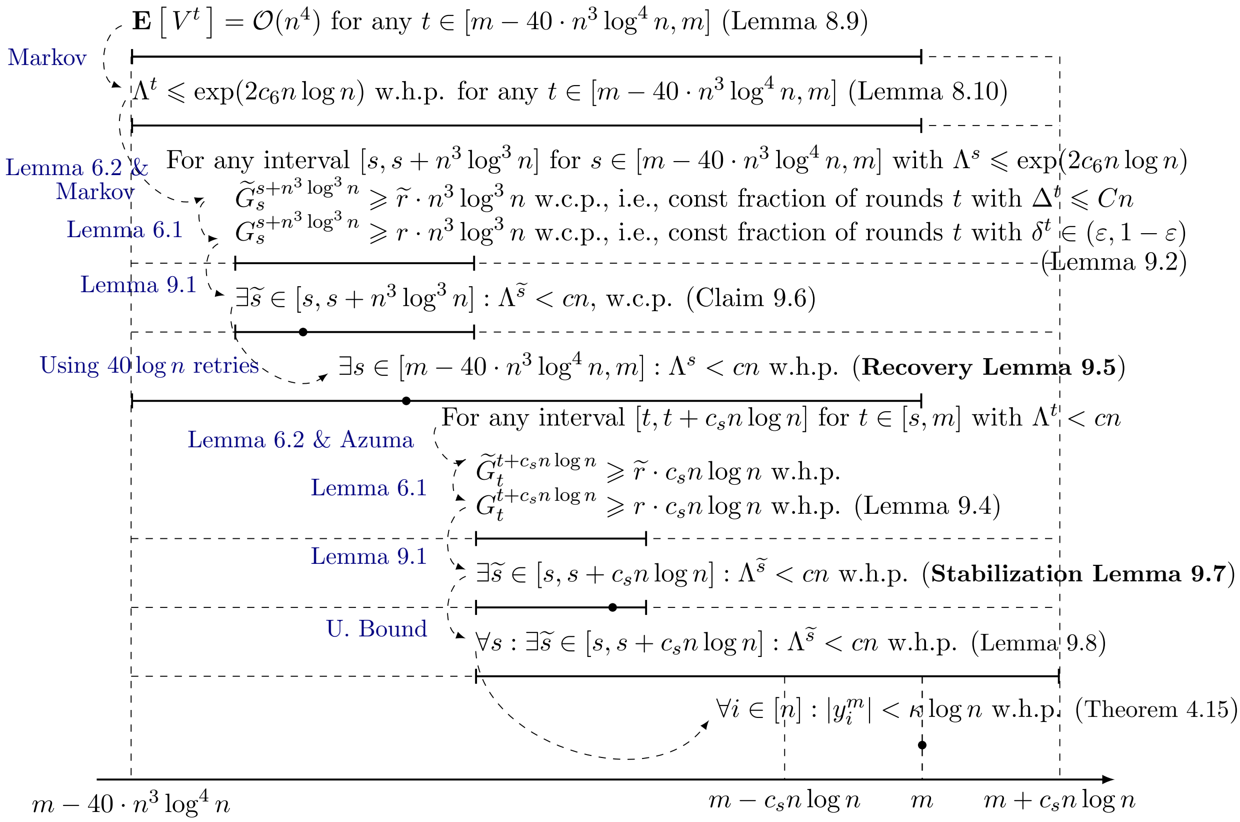

Figure 5: Summary of the key steps in the proof for Theorem 4.15.

There are a considerable number of constants in this paper and their interdependence can be quite complex. We hope the following remark will shed some light on their respective roles.

Remark 6.3 (Relationship Between the Constants). *All the constants used in the analysis depend on the constants w_-, w_+ (and k_1, k_2 if $\mathcal{P}_3$ holds) of the process being analyzed. The relation between the absolute value and quadratic potential functions (and specifically c_1 and c_2 in Lemma 6.2) defines what it means for Δ^t to be small, i.e., $\Delta^t \leqslant C \cdot n$. This C in turn specifies the constant $\varepsilon > 0$, given by Lemma 6.1, which defines a good quantile to be $\delta^t \in (\varepsilon, 1 - \varepsilon)$. Then, ε specifies the fraction r of rounds with $\delta^t \in (\varepsilon, 1 - \varepsilon)$ (in Lemmas 9.3 and 9.4) and the constants c_3, c_4 in the inequalities for the expected change of Λ^t for sufficiently small constant α (in Lemmas 8.4 and 8.5). Then, using the constants ε, r and c_3, c_4 we finally set the value of α in Λ. Finally, ε and α specify c in Corollary 8.6 which we use in proving $\Lambda^t < 2cn$.*

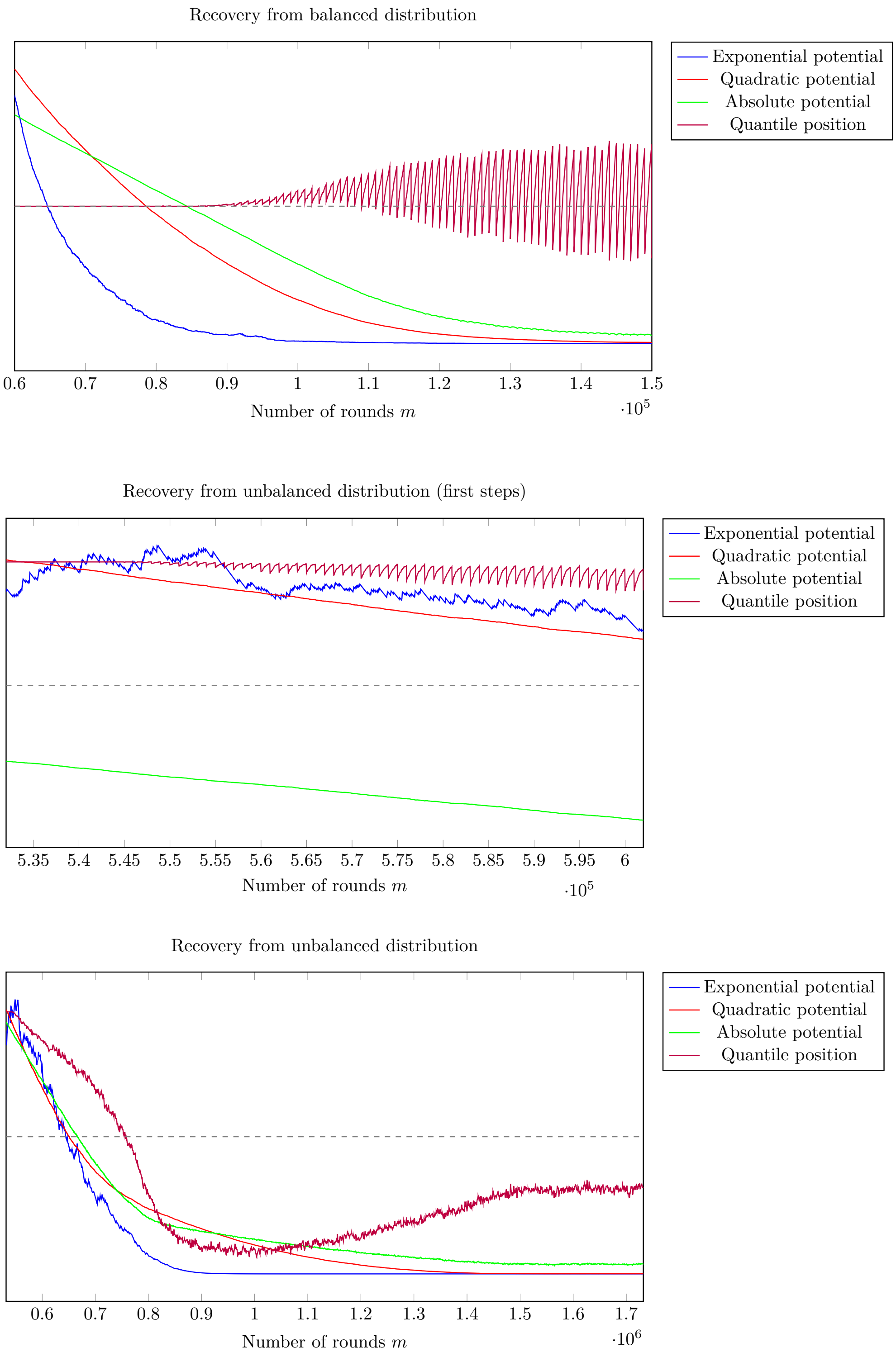

Figure 6: Scaled versions of the potential functions for MEAN-THINNING with different initial load distributions for $n = 1000$. The first example above starts from a load distribution where half of the bins have (normalized) load $+\log n$ and the other half have load $-\log n$. The second example starts from a load distribution such that the quantile of the average load remains very close to 1 for $\omega(n \log n)$ many rounds. As it can be seen, the exponential potential increases a bit at the beginning, but both the absolute and quadratic potential all improve immediately. Once the quantile is sufficiently bounded away from 0 and 1, the exponential potential also decreases, eventually stabilizing at $\mathcal{O}(n)$, as shown in the third figure. Note that the dashed gray line corresponds a perfectly balanced mean quantile, i.e., $\delta^t = 1/2$.

Organization of the Remaining Part of the Proof.

- In Section 7, we prove the mean quantile stabilization lemma (Lemma 6.1) and the relation between the absolute value and the quadratic potential (Lemma 6.2).
- In Section 8, we prove simple relations between the quadratic and the exponential potential (Section 8.1), analyze the expected change of the exponential potential Λ (Section 8.2) and prove that $\mathbf{E}\left[\,V^t\,\right] = \mathrm{poly}(n)$ (Section 8.3), deducing an $\mathcal{O}(n \log n)$ gap.
- In Section 9, we complete the proof for the $\mathcal{O}(\log n)$ gap for non-filling processes by analyzing the recovery (Section 9.3) and stabilization (Section 9.4) phases.

7 Mean Quantile Stabilization

In Section 7.1, we prove the stabilization of the mean quantile δ^t. In Section 7.2, we relate the change of the quadratic potential Υ^t to the absolute value potential Δ^t.

7.1 Mean Quantile Stabilization based on Absolute Value Potential

The next lemma proves that once the absolute value potential is $\mathcal{O}(n)$, then the allocation process will satisfy $\delta^t \in (\varepsilon, 1-\varepsilon)$ for "many" of the following rounds. The condition on δ^t means that a constant fraction of bins are overloaded and a constant fraction are underloaded. Interestingly, this lemma does not need the stronger property $\mathcal{P}_3$ or $\mathcal{W}_3$.

Lemma 6.1 (Mean Quantile Stabilization, restated). *Consider any allocation process satisfying $\mathcal{P}_2$ and $\mathcal{W}_2$. Then for any integer constant $C \geqslant 1$, there exists some $\varepsilon := \varepsilon(C) > 0$ such that for any integers $t_0 \geqslant 0$ and $t_3 := t_0 + \left\lceil \frac{2Cn}{w_+} \right\rceil + \left\lceil \frac{n}{w_+} \right\rceil + \left\lceil \frac{n}{10w_-} \right\rceil$ we have*

$$\mathbf{Pr}\left[\,\left|\left\{t \in [t_0, t_3] \colon \delta^t \in (\varepsilon, 1-\varepsilon)\right\}\right| \geqslant \varepsilon \cdot n \;\middle|\; \mathfrak{F}^{t_0}, \Delta^{t_0} \leqslant C \cdot n\,\right] \geqslant 1 - e^{-\varepsilon \cdot n}.$$

Proof. Define $B_* := \left\{i \in [n] \colon |y_i^{t_0}| < 2C\right\}$ to be the bins whose load deviates from the average W^t/n by less than $2C$. Then,

$$C \cdot n \geqslant \Delta^{t_0} = \sum_{i \in [n]} |y_i^{t_0}| \geqslant \sum_{i \colon |y_i^{t_0}| \geqslant 2C} 2C = (n - |B_*|) \cdot 2C,$$

rearranging this gives that conditional on the event $\mathcal{C} := \{\Delta^{t_0} \leqslant C \cdot n\}$ we have

$$|B_*| \geqslant \frac{n \cdot C}{2C} = \frac{n}{2}. \tag{7.1}$$

Note that the bins in B_* may be underloaded or overloaded.

We now proceed with two claims. Using the fact that $|B_*|$ is large, the first claim (Claim 7.1) proves that there exists an "early" round $t \in [t_0, t_1]$ such that a constant fraction of bins are underloaded. Similarly, the second claim (Claim 7.2) proves that there exists a round $t \in [t_2, t_3]$ with $t_2 \geqslant t_1$ such that a constant fraction of bins are overloaded. Since the set of overloaded bins can only increase by 1 per round, we then finally conclude that for $\Omega(n)$ rounds t, both conditions hold, i.e., $\delta^t \in (\varepsilon, 1-\varepsilon)$.

Claim 7.1. *For any integer constant $C \geqslant 1$, there exists a constant $\kappa_1 := \kappa_1(C) > 0$ such that for $t_1 := t_0 + \left\lceil \frac{2C}{w_+} \cdot n \right\rceil$ we have*

$$\mathbf{Pr}\left[\,\bigcup_{t \in [t_0, t_1]} \left\{\delta^t \geqslant \kappa_1\right\} \;\middle|\; \mathfrak{F}^{t_0}, \Delta^{t_0} \leqslant C \cdot n\,\right] \geqslant 1 - e^{-\kappa_1 \cdot n}.$$

Proof of Claim. If $|B_-^{t_0} \cap B_*| \geqslant 1/2 \cdot |B_*|$, then the statement of the claim follows immediately for $t = t_0$ and $\kappa_1 = 1/4$. Otherwise, assume $|B_+^{t_0} \cap B_*| \geqslant 1/2 \cdot |B_*|$, i.e., at least half of the bins in B_* are overloaded at round t_0. Note that the bins in $B_+^{t_0} \cap B_*$ all have loads in the range $[W^{t_0}/n,\ W^{t_0}/n + 2C)$. Thus, since loads are integers (as w_+ and w_- are integers), there can be at most $2C$ different load levels within $B_+^{t_0} \cap B_*$. Hence, by the pigeonhole principle and (7.1), there exists a subset $\widetilde{B_*} \subseteq B_+^{t_0} \cap B_*$, with $|\widetilde{B_*}| \geqslant |B_+^{t_0} \cap B_*| \cdot \frac{1}{2C} \geqslant \frac{1}{8C} \cdot n$ such that all bins $i \in \widetilde{B_*}$ have the same (non-negative) load at round t_0.

Note that if a bin $i \in \widetilde{B_*}$ was never chosen for an allocation during rounds $[t_0, t_1]$, then its normalized load would satisfy

$$y_i^{t_1} \leqslant y_i^{t_0} - \frac{w_+}{n} \cdot (t_1 - t_0) \leqslant y_i^{t_0} - 2C < 2C - 2C = 0,$$

and hence the bin would become underloaded at least once before round $t_1 = t_0 + \lceil 2Cn/w_+ \rceil$.

Since for any round, a fixed overloaded bin is chosen for an allocation with probability at most $p_+^t \leqslant \frac{1}{n}$, we conclude for any $i \in \widetilde{B_*}$,

$$\mathbf{Pr}\left[i \in \bigcup_{t \in [t_0, t_1]} B_-^t \;\Big|\; \mathfrak{F}^{t_0} \right] \geqslant \mathbf{Pr}\left[\mathsf{Bin}\left(\frac{2C}{w_+} \cdot n, \frac{1}{n} \right) = 0 \right] = \left(1 - \frac{1}{n} \right)^{n \cdot 2C/w_+} \geqslant 4^{-2C/w_+},$$

where in the last inequality we used that $\left(1 - \frac{1}{n}\right)^n$ is non-decreasing in $n \geqslant 2$. Define the constant $\kappa_2 := 4^{-2C/w_+} > 0$. Further, let

$$Y := \left| \bigcup_{t \in [t_0, t_1]} B_-^t \cap \widetilde{B_*} \right|,$$

be the number of bins in $\widetilde{B_*}$ that become underloaded at least once during the interval $[t_0, t_1]$.

Let us now consider a modified process in which the probability for each overloaded bin to be chosen is not only at most $\frac{1}{n}$, but is instead equal to $\frac{1}{n}$ in each round. Let $\widetilde{Y}$ be the number of bins in $\widetilde{B_*}$ that are never chosen in the modified process during the interval $[t_0, t_1] = [t_0, t_0 + \lceil 2Cn/w_+ \rceil]$. Note that $\widetilde{Y}$ is stochastically smaller than Y.

Recall that $\mathcal{C} = \{\Delta^{t_0} \leqslant C \cdot n\}$. By linearity of expectation and (7.1), we have

$$\mathbf{E}[\, \widetilde{Y} \mid \mathfrak{F}^{t_0},\, \mathcal{C} \,] \geqslant \kappa_2 \cdot \mathbf{E}\left[|\widetilde{B_*}| \mid \mathfrak{F}^{t_0},\, \mathcal{C} \right] \geqslant \kappa_2 \cdot \frac{1}{4C} \mathbf{E}\left[|B_*| \mid \mathfrak{F}^{t_0},\, \mathcal{C} \right] \geqslant \kappa_2 \cdot \frac{1}{4C} \cdot \frac{n}{2} = \kappa_2 \cdot \frac{n}{8C}.$$

In the modified process, changing the bin sample in one round can change $\widetilde{Y}$ only by at most one, and hence applying the Method of Bounded Independent Differences (Lemma A.8) yields,

$$\mathbf{Pr}\left[\widetilde{Y} \leqslant \frac{1}{2} \mathbf{E}[\, \widetilde{Y} \mid \mathfrak{F}^{t_0},\, \mathcal{C} \,] \;\Big|\; \mathfrak{F}^{t_0},\, \mathcal{C} \right] \leqslant \exp\left(- \frac{\frac{1}{4}\left(\mathbf{E}[\, \widetilde{Y} \mid \mathfrak{F}^{t_0},\, \mathcal{C} \,] \right)^2}{\frac{2C}{w_+} \cdot n \cdot 1^2} \right) \leqslant \exp\left(- \frac{\kappa_2^2 w_+}{8^3 C^3} \cdot n \right),$$

and so by stochastic domination and the two equations above we have

$$\mathbf{Pr}\left[Y \geqslant \frac{\kappa_2 n}{16C} \;\Big|\; \mathfrak{F}^{t_0},\, \mathcal{C} \right] \geqslant 1 - \exp\left(- \frac{\kappa_2^2 w_+}{8^3 C^3} \cdot n \right).$$

Since all load levels in $\widetilde{B_*}$ before round t_0 are identical, we conclude that they all become underloaded the first time in the same round, so

$$\mathbf{Pr}\left[|B_-^t \cap \widetilde{B_*}| \geqslant \kappa_1 \cdot n \mid \mathfrak{F}^{t_0},\, \Delta^{t_0} \leqslant C \cdot n \right] \geqslant 1 - e^{-\kappa_1 \cdot n},$$

where $\kappa_1 := \min\left\{ \frac{\kappa_2}{16C}, \frac{\kappa_2^2 w_+}{8^3 C^3} \right\} > 0$ and $\kappa_2 = 4^{-2C/w_+}$, as defined above. $\Diamond$

Claim 7.2. *For any integer constant $C \geqslant 1$ there exists a constant $\kappa_3 := \kappa_3(C) > 0$ such that for $t_2 := t_1 + \lceil \frac{n}{w^+} \rceil$, and $t_3 := t_2 + \left\lceil \frac{n}{10w_-} \right\rceil$ we have*

$$\mathbf{Pr}\left[\bigcup_{t\in[t_2,t_3]} \{\delta^t \leqslant 1-\kappa_3\} \,\middle|\, \mathfrak{F}^{t_0}, \Delta^{t_0} \leqslant C\cdot n \right] \geqslant 1 - e^{-\kappa_3 \cdot n}.$$

Proof of Claim. If $|B_* \cap B_+^{t_2}| \geqslant \frac{1}{2}\cdot|B_*| \geqslant \frac{n}{4}$, the claim follows immediately, so assume $|B_* \cap B_-^{t_2}| \geqslant \frac{1}{2}\cdot|B_*|$. Let $\widetilde{B_*} := B_* \cap B_-^{t_2}$ and consider any bin i in $\widetilde{B_*}$. After round t_2, the normalized load satisfies

$$y_i^{t_2} \geqslant y_i^{t_0} - (t_2 - t_0)\cdot \frac{w_-}{n} \geqslant -2C - \left(\frac{2C+1}{w_+} + \frac{2}{n}\right)\cdot w_- \geqslant -\left(5C + \frac{2}{n}\right)\cdot w_-,$$

where we used $y_i^{t_0} > -2C$ as $i \in \widetilde{B_*} \subseteq B_*$, $C \geqslant 1$, and $w_- \geqslant w_+ \geqslant 1$ by $\mathcal{W}_2$.

As long as a bin i is underloaded, we choose it for allocation in each round with probability at least $p_-^t \geqslant 1/n$, independently of previous rounds. If the bin has been chosen $6C$ times (while being underloaded), it must become overloaded before round t_3 at least once since

$$y_i^{t_3} \geqslant y_i^{t_2} - (t_3 - t_2)\frac{w_-}{n} + 6C\cdot w_- \geqslant -\left(5C + \frac{2}{n}\right)\cdot w_- - \left(\frac{1}{10} + \frac{w_-}{n}\right) + 6C\cdot w_- > 0.$$

Note that the lower bound of $1/n$ on the probability for allocating to any underloaded bin in a single round t holds in any round, regardless of the ball configuration. Hence, for any $i \in \widetilde{B_*}$,

$$\begin{aligned}\mathbf{Pr}\left[i \in \bigcup_{t\in[t_2,t_3]} B_+^t \,\middle|\, \mathfrak{F}^{t_0} \right] &\geqslant \mathbf{Pr}\left[\mathsf{Bin}\left(\frac{n}{10w_-}, \frac{1}{n}\right) \geqslant 6C \right] \\ &\geqslant \binom{\frac{n}{10w_-}}{6C}\left(\frac{1}{n}\right)^{6C}\cdot\left(1-\frac{1}{n}\right)^{\frac{n}{10w_-}-6C} \\ &\geqslant \left(\frac{1}{60Cw_-}\right)^{6C}\cdot e^{-\frac{1}{10w_-}}/2,\end{aligned}$$

where we used Lemma A.11 and $\binom{n}{k} \geqslant (\frac{n}{k})^k$ in the last inequality. Define the constant $\kappa_4 := (60Cw_-)^{-6C}\cdot e^{-\frac{1}{10w_-}}/2 > 0$ and let

$$Z := \left| \bigcup_{t\in[t_2,t_3]} B_+^t \cap \widetilde{B_*} \right|,$$

that is, the number of bins in $\widetilde{B_*}$ that become overloaded at least once during the interval $[t_2, t_3]$. Similar to the proof of the previous claim, consider a modified process where each underloaded bin has a probability of exactly $\frac{1}{n}$ to be incremented in each round. Let $\widetilde{Z}$ be the number of bins in $\widetilde{B_*}$, where $|\widetilde{B_*}| \geqslant n/4$ by (7.1) if we condition on $\mathcal{C} = \Delta^{t_0} \leqslant C\cdot n$, that are chosen at least $6C$ times in this modified process during rounds $[t_2, t_3]$. Note that for any $\mathfrak{F}^{t_0}$, $\widetilde{Z}$ is stochastically smaller than Z. By linearity of expectations we have

$$\mathbf{E}[\,\widetilde{Z} \mid \mathfrak{F}^{t_0},\, \mathcal{C}\,] \geqslant \kappa_4 \cdot \mathbf{E}\left[|\widetilde{B_*}| \,\middle|\, \mathfrak{F}^{t_0},\, \mathcal{C} \right] \geqslant \kappa_2 \cdot \frac{1}{2}\cdot \mathbf{E}\left[\,|B_*| \mid \mathfrak{F}^{t_0},\, \mathcal{C}\,\right] \geqslant \kappa_2\cdot\frac{1}{2}\cdot\frac{n}{2} = \kappa_2\cdot\frac{n}{4}.$$

In the modified process, changing the bin sample in one round can change $\widetilde{Z}$ only by at most one, and hence applying the Method of Bounded Independent Differences (Lemma A.8) yields,

$$\mathbf{Pr}\left[\widetilde{Z} < \frac{1}{2}\cdot \mathbf{E}[\,\widetilde{Z} \mid \mathfrak{F}^{t_0},\, \mathcal{C}\,] \,\middle|\, \mathfrak{F}^{t_0},\, \mathcal{C} \right] \leqslant \exp\left(-\frac{\frac{1}{4}\left(\mathbf{E}[\,\widetilde{Z} \mid \mathfrak{F}^{t_0},\, \mathcal{C}\,]\right)^2}{2\cdot(t_3-t_2)\cdot 1^2} \right) = \exp\left(-\frac{5\kappa_4^2 \cdot w_-}{2^4}\cdot n \right).$$

and by the stochastic domination, we have

$$\mathbf{Pr}\left[\, Z \geqslant \frac{\kappa_4}{4} \cdot n \,\middle|\, \mathfrak{F}^{t_0},\, \mathcal{C} \,\right] \geqslant 1 - \exp\left(-\frac{5\kappa_4^2 \cdot w_-}{2^4} \cdot n \right).$$

We now fix the constant $\kappa_3 := \min\left\{ \frac{1}{2} \cdot \frac{\kappa_4}{4},\ \frac{5\kappa_4^2 \cdot w_-}{2^4} \right\}$, where $\kappa_4 > 0$ is given above.

At this point we see that with probability at least $1 - e^{-\kappa_3 n}$ at least $2\kappa_3 \cdot n$ bins become overloaded at least once at some round in $[t_2, t_3]$. However, for each bin, there is the possibility that it is overloaded only for one round before becoming underloaded again. If each bin is overloaded for only one round, and all at different rounds, then it is possible that at no single round in $[t_2, t_3]$ do we have $\Omega(n)$ overloaded bins. We now explain why this is not the case.

Note that during the interval $[t_2, t_3]$, as $t_3 - t_2 = \left\lceil \frac{n}{10w_-} \right\rceil$, there is at most one round t where the integer parts of the average load changes, that is, there is only at most one round $t \in [t_2, t_3)$ such that $\lfloor W^t/n \rfloor < \lfloor W^{t+1}/n \rfloor$. Hence only at the transition from t to $t+1$ could an overloaded bin become underloaded during the interval $[t_2, t_3]$. Therefore, if $|Z| \geqslant 2\kappa_3 \cdot n$, we must have

$$|B_+^t \cap B_*| \geqslant \kappa_3 \cdot n \qquad \text{or} \qquad |B_+^{t_3} \cap B_*| \geqslant \kappa_3 \cdot n.$$

Hence $|Z| \geqslant 2\kappa_3 \cdot n$, implies that there exists $t \in [t_2, t_3]$ such that $|B_+^t| \geqslant \kappa_3 \cdot n$. $\Diamond$

The first claim implies that, w.p. $1 - e^{-\kappa_1 n}$, that the process reaches a round $s_1 \in [t_0, t_1]$ with $\kappa_1 \cdot n$ underloaded bins. The second claim shows, w.p. $1 - e^{-\kappa_3 n}$, that the process reaches a round $s_2 \in [t_1 + \lceil \frac{n}{w_+} \rceil, t_3]$ with $\kappa_3 \cdot n$ overloaded bins. By the union bound, both events occur with probability $1 - e^{-\kappa_1 n} - e^{-\kappa_3 n}$. Since $\delta^t = |B_+^t|/n$, we have $\delta^{t+1} \leqslant \delta^t + \frac{1}{n}$, and in this case applying Lemma A.7 with $r_0 = s_1$, $r_1 = s_2$, $f(t) = \delta^t$, $\varepsilon := \min\{\kappa_3, \kappa_1, 2/3\}$ and $\xi = 1/n$ gives,

$$\begin{aligned}
\left|\left\{ t \in [t_0, t_3] \colon \delta^t \in (\varepsilon/2, 1 - \varepsilon/2) \right\}\right| &\geqslant \left|\left\{ t \in [s_1, s_2] \colon \delta^t \in (\varepsilon/2, 1 - \varepsilon/2) \right\}\right| \\
&\geqslant \min\{\varepsilon/2 \cdot n, s_2 - s_1\} \\
&\geqslant \min\{\varepsilon/2 \cdot n, \frac{n}{w^+}\}.
\end{aligned}$$

Thus, taking ε in the statement to be $\min\{\min\{\kappa_3, \kappa_1, 2/3\}/2, 1/w_+\}$ gives the result. $\square$

7.2 Quadratic and Absolute Value Potential Functions

Lemma 6.2 (restated). *Consider any allocation process satisfying $\mathcal{W}_2$ and $\mathcal{P}_2$. Then for any $t \geqslant 0$, the quadratic potential satisfies*

$$\mathbf{E}\left[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - (p_-^t \cdot w_- - p_+^t \cdot w_+) \cdot \Delta^t + 4 \cdot (w_-)^2.$$

Hence for any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process, this implies by Lemma 4.14 that there exist constants $c_1, c_2 > 0$ such that for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - \frac{c_1}{n} \cdot \Delta^t + c_2.$$

Proof. We begin by decomposing the potential Υ over the n bins, using $\Upsilon_i^{t+1} := (y_i^{t+1})^2$:

$$\Upsilon^{t+1} = \sum_{i=1}^n \Upsilon_i^{t+1} = \sum_{i=1}^n (y_i^{t+1})^2.$$

We shall analyze each term Υ_i^{t+1} separately in two cases depending on the load of the bin i. By the $\mathcal{W}_2$ assumption, if the chosen bin is overloaded, we place a ball of weight w_+, otherwise we place a ball of weight w_-. Also, recall that $P_+^t := \sum_{i \in B_+^t} p_i^t$ and $P_-^t := \sum_{i \in B_-^t} p_i^t$.

Case 1 $[i \in B_+^t]$. If i is assigned to an overloaded bin,

$$\mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] = \underbrace{p_i^t \cdot \Big(y_i^t + w_+ - \frac{w_+}{n}\Big)^2}_{\text{placing a ball in } i} + \underbrace{(P_+^t - p_i^t) \cdot \Big(y_i^t - \frac{w_+}{n}\Big)^2}_{\text{placing a ball in } B_+^t \setminus \{i\}} + \underbrace{P_-^t \cdot \Big(y_i^t - \frac{w_-}{n}\Big)^2}_{\text{placing a ball in } B_-^t}.$$

Expanding out the squares,

$$\begin{aligned}
\mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] &= (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot \Big(w_+ - \frac{w_+}{n}\Big) - (P_+^t - p_i^t) \cdot \frac{w_+}{n} - P_-^t \cdot \frac{w_-}{n}\Big) \\
&\quad + p_i^t \cdot \Big(w_+ - \frac{w_+}{n}\Big)^2 + (P_+^t - p_i^t) \cdot \frac{(w_+)^2}{n^2} + P_-^t \cdot \frac{(w_-)^2}{n^2} \\
&= (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot w_+ - P_+^t \cdot \frac{w_+}{n} - P_-^t \cdot \frac{w_-}{n}\Big) \\
&\quad + p_i^t \cdot (w_+)^2 - 2 \cdot p_i^t \cdot \frac{(w_+)^2}{n} + P_+^t \cdot \frac{(w_+)^2}{n^2} + P_-^t \cdot \frac{(w_-)^2}{n^2} \\
&\leqslant (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot w_+ - P_+^t \cdot \frac{w_+}{n} - P_-^t \cdot \frac{w_-}{n}\Big) + 2 \cdot \frac{(w_-)^2}{n},
\end{aligned}$$

where in the last step we used that $p_i^t \leqslant p_+^t \leqslant 1/n$ by $\mathcal{P}_2$, which implies $p_i^t \cdot (w_+)^2 + P_+^t \cdot \frac{(w_+)^2}{n^2} + P_-^t \cdot \frac{(w_-)^2}{n^2} \leqslant 2 \cdot \frac{(w_-)^2}{n}$.

Case 2 $[i \in B_-^t]$. If i is assigned to an underloaded bin then

$$\mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] = \underbrace{p_i^t \cdot \Big(y_i^t + w_- - \frac{w_-}{n}\Big)^2}_{\text{placing a ball in } i} + \underbrace{(P_-^t - p_i^t) \cdot \Big(y_i^t - \frac{w_-}{n}\Big)^2}_{\text{placing a ball in } i \in B_-^t \setminus \{i\}} + \underbrace{P_+^t \cdot \Big(y_i^t - \frac{w_+}{n}\Big)^2}_{\text{placing a ball in } i \in B_+^t}.$$

Expanding out the squares,

$$\begin{aligned}
\mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] &= (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot \Big(w_- - \frac{w_-}{n}\Big) - (P_-^t - p_i^t) \cdot \frac{w_-}{n} - P_+^t \cdot \frac{w_+}{n}\Big) \\
&\quad + p_i^t \cdot \Big(w_- - \frac{w_-}{n}\Big)^2 + (P_-^t - p_i^t) \cdot \frac{(w_-)^2}{n^2} + P_+^t \cdot \frac{(w_+)^2}{n^2} \\
&= (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot w_- - P_-^t \cdot \frac{w_-}{n} - P_+^t \cdot \frac{w_+}{n}\Big) \\
&\quad + p_i^t \cdot (w_-)^2 - 2 \cdot p_i^t \cdot \frac{(w_-)^2}{n} + P_-^t \cdot \frac{(w_-)^2}{n^2} + P_+^t \cdot \frac{(w_+)^2}{n^2} \\
&\leqslant (y_i^t)^2 + 2y_i^t \cdot \Big(p_i^t \cdot w_- - P_-^t \cdot \frac{w_-}{n} - P_+^t \cdot \frac{w_+}{n}\Big) + 2 \cdot p_i^t \cdot (w_-)^2,
\end{aligned}$$

where in the last step we used that $p_i^t \geqslant p_-^t \geqslant 1/n$ by $\mathcal{P}_2$, which implies $p_i^t \cdot (w_-)^2 + P_-^t \cdot \frac{(w_-)^2}{n^2} + P_+^t \cdot \frac{(w_+)^2}{n^2} \leqslant 2 \cdot p_i^t \cdot (w_-)^2$.

Combining the two inequalities for the two cases and since $\sum_{i \in B_+^t} y_i^t = -\sum_{i \in B_-^t} y_i^t = \frac{1}{2} \cdot \Delta^t$ we get,

$$\begin{aligned}
\mathbf{E}\left[\,\Upsilon^{t+1} \mid \mathfrak{F}^t\,\right] &= \sum_{i \in B_+^t} \mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] + \sum_{i \in B_-^t} \mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] \\
&\leqslant \Upsilon^t + \sum_{i \in B_+^t} 2y_i^t p_i^t \cdot w_+ + \sum_{i \in B_-^t} 2y_i^t p_i^t \cdot w_- + \sum_{i \in B_+^t} 2 \cdot \frac{(w_-)^2}{n} + \sum_{i \in B_-^t} 2 \cdot p_i^t \cdot (w_-)^2 \\
&\leqslant \Upsilon^t - (p_-^t \cdot w_- - p_+^t \cdot w_+) \cdot \Delta^t + 4 \cdot (w_-)^2,
\end{aligned}$$

where the last inequality follows since $p_i^t \leqslant p_+^t$ for $i \in B_+^t$, and $p_i^t \geqslant p_-^t$ for $i \in B_-^t$. $\square$

8 Potential Function Inequalities

In this section we derive several inequalities involving potential functions. Most of the effort goes into establishing a drop of the exponential potential function for some suitable choices of α, which in turn depends on the constants defined by the process. One of the main insights is Corollary 8.6, which establishes: (i) a significant drop of the exponential potential function if the quantile satisfies $\delta^t \in (\varepsilon, 1-\varepsilon)$, and (ii) a not too large increase of the exponential potential function for any quantile.

On a high level the analysis follows relatively standard estimates and bears resemblance to the one in [24]. Even though some extra care is needed due to the more general allocation process, a reader may wish to skip this section (or the proofs) and continue with the proof in Section 9.

8.1 Quadratic and Exponential Potential Functions

The next lemma bounds the quadratic potential in terms of the exponential potential.

Lemma 8.1. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. For any $0 < \alpha < 1$. Then for any $t \geqslant 0$,*

$$\Upsilon^t \leqslant \alpha^{-2} \cdot n \cdot (\log \Lambda^t)^2, \qquad \text{and} \qquad \Upsilon^t \leqslant \left(\frac{4}{\alpha} \cdot \log\left(\frac{4}{\alpha}\right)\right)^2 \cdot \Lambda^t.$$

Proof. Note that by the definition of Λ^t, if $\Lambda^t \leqslant \lambda$ then $y_i^t \leqslant \frac{1}{\alpha} \cdot \log \lambda$ and $-y_i^t \leqslant \frac{1}{\alpha} \cdot \log \lambda$ for all $i \in [n]$. Hence, $(y_i^t)^2 \leqslant \frac{1}{\alpha^2} \cdot (\log \lambda)^2$ for all $i \in [n]$, which proves the first statement by aggregating over all bins.

For the second statement let $\kappa := ((4/\alpha) \cdot \log(4/\alpha))^2$. Note that $e^y \geqslant y$ (for any $y \geqslant 0$) and hence for any $y \geqslant (4/\alpha) \cdot \log(4/\alpha)$,

$$e^{\alpha y/2} = e^{\alpha y/4} \cdot e^{\alpha y/4} \geqslant \frac{\alpha y}{4} \cdot \frac{4}{\alpha} \geqslant y.$$

Hence for $y \geqslant (4/\alpha) \cdot \log(4/\alpha)$,

$$e^{\alpha y} = e^{\alpha y/2} \cdot e^{\alpha y/2} \geqslant y \cdot y = y^2.$$

Thus we conclude

$$\Upsilon^t = \sum_{i=1}^n (y_i^t)^2 \leqslant \sum_{i=1}^n \max\left\{\Lambda_i^t, \kappa\right\} \leqslant \sum_{i=1}^n \max\left\{\Lambda_i^t, \Lambda_i^t \kappa\right\} \leqslant \Lambda^t \cdot \kappa,$$

where the third inequality used that $\Lambda_i^t \geqslant 1$ for any $i \in [n]$ and the fourth used $\kappa \geqslant 1$. $\square$

The next lemma is very basic but is used in Lemma 9.4 so we prove it for completeness.

Lemma 8.2. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. For any $t \geqslant 0$ we have*

$$|\Upsilon^{t+1} - \Upsilon^t| \leqslant 4w_- \cdot \max_{i \in [n]} |y_i^t| + 2w_-^2.$$

Proof. Recall that $w^t = W^{t+1} - W^t$ is the total number of balls allocated in the round t. To begin since only one bin, we shall call this i, is updated at each round we have

$$\begin{aligned} |\Upsilon^{t-1} - \Upsilon^t| &= \left|\sum_{j=1}^n (y_j^{t+1})^2 - \sum_{j=1}^n (y_j^t)^2\right| \\ &\leqslant \left|\left(y_i^t + w^t \cdot \left(1 - \frac{1}{n}\right)\right)^2 - (y_i^t)^2\right| + \sum_{j \in [n], j \neq i} \left|\left(y_j^t - \frac{w^t}{n}\right)^2 - (y_j^t)^2\right|. \end{aligned}$$

Now for any $j \in [n]$ we have

$$\left| \left(y_j^t - \frac{w^t}{n}\right)^2 - \left(y_j^t\right)^2 \right| \leqslant 2 \left|\frac{w^t}{n}\right| \cdot |y_i^t| + \left|\frac{w^t}{n}\right|^2 \leqslant \frac{2w_-}{n} \cdot \max_{i \in [n]} |y_i^t| + \frac{w_-^2}{n^2},$$

since $W^t \leqslant W^{t+1} \leqslant W^t + w_-$ and $n \geqslant 1$. Similarly for any $i \in [n]$ we have

$$\begin{aligned}
&\left| \left(y_i^t + w^t \cdot \left(1 - \frac{1}{n}\right)\right)^2 - \left(y_i^t\right)^2 \right| \\
&\leqslant 2 \cdot |y_i^t| \cdot \left| w^t \cdot \left(1 - \frac{1}{n}\right)\right| + \left| w^t \cdot \left(1 - \frac{1}{n}\right)\right|^2 \\
&\leqslant 2 \max_{i \in [n]} |y_i^t| \cdot w_- + w_-^2. \qquad \square
\end{aligned}$$

8.2 Exponential Potential Λ

In this section, we consider the exponential potential Λ. Let $\mathcal{G}^t$ be the event that $\delta^t \in (\varepsilon, 1-\varepsilon)$ holds. We will prove in Lemma 8.4 that the potential drops in expectation when the quantile is good after round t, i.e.,

$$\mathbf{E}[\, \Lambda^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t \,] \leqslant \Lambda^t \cdot \left(1 - \frac{2c_3\alpha}{n}\right) + c_3',$$

and in Lemma 8.5 that it has a bounded increase at a round when the quantile is not good,

$$\mathbf{E}[\, \Lambda^{t+1} \mid \mathfrak{F}^t, \neg\mathcal{G}^t \,] \leqslant \Lambda^t \cdot \left(1 + \frac{c_4\alpha^2}{2n}\right) + c_4.$$

Note that the decrease factor can be made arbitrarily larger than the increase factor, by choosing $\alpha > 0$ smaller. So, once we prove that there is a constant fraction of rounds with a good quantile (Lemma 9.4), we can deduce that there is overall an expected decrease in the exponential potential (Lemma 9.1 and Lemma 9.7), when the potential is sufficiently large. Note that the exponential potential cannot decrease in expectation in every round (Claim B.2).

For the analysis of Λ as well as V (in Section 8.3), we consider the labeling of the bins $i \in [n]$ used by the allocation process in round t so that x_i^t is non-decreasing in $i \in [n]$. We write

$$\Lambda^t =: \sum_{i=1}^n \Lambda_i^t = \sum_{i=1}^n e^{\alpha |y_i^t|} \quad \left(\text{and} \quad V^t =: \sum_{i=1}^n V_i^t = \sum_{i=1}^n e^{\widetilde{\alpha} |y_i^t|}\right),$$

and handle separately the following three cases of bins based on their load:

- **Case 1** [Robustly Overloaded Bins]. The set of bins B_{++}^t with load $y_i^t \geqslant \frac{w_-}{n}$. These are bins in B_+^t that are guaranteed to be in B_+^{t+1} (that is, overloaded), since the average load can increase by at most w_-/n.

  For the exponential potential Λ^t or (and V^t respectively), the change of a single bin $i \in B_{++}^t$ is given by,

$$\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] = \Lambda_i^t \cdot \Big(\underbrace{p_i^t \cdot e^{\alpha w_+ - \alpha w_+/n}}_{\text{placing a ball in } i} + \underbrace{(P_+^t - p_i^t) \cdot e^{-\alpha w_+/n}}_{\text{placing a ball in } B_+^t \setminus \{i\}} + \underbrace{P_-^t \cdot e^{-\alpha w_-/n}}_{\text{placing a ball in } B_-^t} \Big).$$

- **Case 2** [Robustly Underloaded Bins]. The set of bins B_{--}^t with load $y_i^t \leqslant -w_-$. These are bins in B_-^t that are guaranteed to be in B_-^{t+1} (that is, underloaded), since a bin can receive a weight of at most w_- in one round.

For the exponential potential Λ^t or (and V^t respectively), the change of a single bin $i \in B_{--}^t$ is given by,

$$\mathbf{E}\left[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,\right] = \Lambda_i^t \cdot \Big(\underbrace{p_i^t \cdot e^{-\alpha w_- + \alpha w_-/n}}_{\text{placing a ball in } i} + \underbrace{(P_-^t - p_i^t) \cdot e^{\alpha w_-/n}}_{\text{placing a ball in } B_-^t \setminus \{i\}} + \underbrace{P_+^t \cdot e^{\alpha w_+/n}}_{\text{placing a ball in } B_+^t} \Big).$$

- **Case 3** [Swinging Bins]. The set of bins $B_{+/-}^t$ with load $y_i^t \in (-w_-, \frac{w_-}{n})$.

We begin by showing that the aggregated contribution of the swinging bins to the change of the potential Λ is at most a constant. This will be used in the proofs of Lemmas 8.4 and 8.5.

Lemma 8.3. *For any constant $\alpha \in (0, 1]$, for any constant $\kappa_1 \geqslant 0$ and any $t \geqslant 0$, we have*

$$\sum_{i \in B_{+/-}^t} \mathbf{E}[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,] \leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \Big(1 - \frac{2\kappa_1\alpha}{n}\Big) + 3(\kappa_1 + w_-) \cdot e^{2w_-}.$$

Proof. For each bin $i \in B_{+/-}^t$, the two events affecting the contribution of the bin to Λ^{t+1} are (i) "internal" due to a ball being placed i and (ii) "external" due to the change in the average.

For (i), the chosen bin $i \in B_{+/-}^t$ can increase by at most w_-, so the potential value satisfies $\Lambda_i^{t+1} \leqslant e^{2\alpha w_-}$. For (ii), the maximum change in the average load is at most w_-/n, and this leads to $\Lambda_i^{t+1} \leqslant \Lambda_i^t \cdot e^{\alpha w_-/n}$. Combining the two contributions

$$\begin{aligned}
\sum_{i \in B_{+/-}^t} \mathbf{E}[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,] &\leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot e^{\alpha w_-/n} \cdot (1 - p_i^t) + \sum_{i \in B_{+/-}^t} e^{2\alpha w_-} \cdot p_i^t \\
&\leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot e^{\alpha w_-/n} + e^{2\alpha w_-} \\
&\overset{(a)}{\leqslant} \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \Big(1 + \frac{2w_-\alpha}{n}\Big) + e^{2\alpha w_-} \\
&\overset{(b)}{\leqslant} \sum_{i \in B_{+/-}^t} \Lambda_i^t + 2\alpha w_- \cdot e^{\alpha w_-} + e^{2\alpha w_-},
\end{aligned}$$

where we used the Taylor estimate $e^z \leqslant 1 + 2z$ for any $z < 1.2$, for sufficiently large n in (a), and $\Lambda_i^t \leqslant e^{\alpha w_-}$ for any $i \in B_{+/-}^t$ in (b). By adding and subtracting $\sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \frac{2\alpha\kappa_1}{n}$,

$$\begin{aligned}
\sum_{i \in B_{+/-}^t} \mathbf{E}[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,] &\leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \Big(1 - \frac{2\alpha\kappa_1}{n}\Big) + \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \frac{2\alpha\kappa_1}{n} + 2\alpha w_- \cdot e^{\alpha w_-} + e^{2\alpha w_-} \\
&\leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \Big(1 - \frac{2\alpha\kappa_1}{n}\Big) + (2\alpha\kappa_1 + 2\alpha w_-) \cdot e^{\alpha w_-} + e^{2\alpha w_-},
\end{aligned}$$

where we used that $\Lambda_i^t \leqslant e^{\alpha w_-}$ and $|B_{+/-}^t| \leqslant n$. Finally we have $(2\alpha\kappa_1 + 2\alpha w_-) \cdot e^{\alpha w_-} + e^{2\alpha w_-} \leqslant (2\kappa_1 + 2w_-) \cdot e^{w_-} + e^{2w_-} \leqslant 3(\kappa_1 + w_-) \cdot e^{2w_-}$, as $w_- \geqslant 1$ and $\alpha \leqslant 1$. ☐

Now, we show that if the quantile δ^t of the mean is in $(\varepsilon, 1 - \varepsilon)$, then the potential function exhibits a multiplicative drop.

Lemma 8.4. *For any $\mathcal{W}_3 \cap \mathcal{P}_3$-process and any constant $\varepsilon \in (0, 1)$, choose a constant $\alpha := \alpha(\varepsilon)$ such that*

$$0 < \alpha \leqslant \min\left\{ \frac{1}{w_-}, \frac{k_2\varepsilon}{2w_-(1 + k_2\varepsilon)}, \frac{k_1\varepsilon}{2w_+(1 - k_1\varepsilon)} \right\}. \tag{8.1}$$

For any $\mathcal{P}_2 \cap \mathcal{W}_2$-process and any constant $\varepsilon \in (0,1)$, choose a constant $\alpha := \alpha(\varepsilon)$ such that

$$0 < \alpha \leqslant \min \left\{ \frac{1}{w_-}, \frac{\varepsilon(w_- - w_+)}{4w_-^2}, \frac{\varepsilon}{2 \cdot w_- \cdot (2+\varepsilon)} \right\}. \tag{8.2}$$

Then there exists a constant $c_3 := c_3(\varepsilon) > 0$ such that for $c_3' := 3(c_3 + w_-) \cdot e^{2w_-}$ and any $t \geqslant 0$ we have

$$\mathbf{E}\left[\Lambda^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t \right] \leqslant \Lambda^t \cdot \left(1 - \frac{2c_3\alpha}{n}\right) + c_3'.$$

Proof. This proof will work with the labeling of the bins $i \in [n]$ used by the allocation process in round t so that x_i^t is non-decreasing in $i \in [n]$.

Case 1 [Robustly Overloaded Bins]. For $i \in B_{++}^t$ in this case

$$\mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] = \Lambda_i^t \cdot \left(p_i^t \cdot e^{-\alpha w_+/n + \alpha w_+} + (P_+^t - p_i^t) \cdot e^{-\alpha w_+/n} + P_-^t \cdot e^{-\alpha w_-/n} \right).$$

Applying the Taylor estimate $e^z \leqslant 1 + z + z^2$, which holds for any $z \leqslant 1.75$, since $\alpha w_+ \leqslant 1$ (and $\alpha w_- \leqslant 1$),

$$\begin{aligned} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] \leqslant \Lambda_i^t \cdot \Bigg(1 &+ p_i^t \cdot \left(-\frac{\alpha w_+}{n} + \alpha w_+ + \left(-\frac{\alpha w_+}{n} + \alpha w_+ \right)^2 \right) \\ &+ (P_+^t - p_i^t) \cdot \left(-\frac{\alpha w_+}{n} + \left(\frac{\alpha w_+}{n} \right)^2 \right) + P_-^t \cdot \left(-\frac{\alpha w_-}{n} + \left(\frac{\alpha w_-}{n} \right)^2 \right) \Bigg). \end{aligned}$$

By gathering $o(n^{-1})$ terms and rearranging terms we obtain,

$$\begin{aligned} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant \Lambda_i^t \cdot \Bigg(1 + p_i^t \cdot \left(-\frac{\alpha w_+}{n} + \alpha w_+ + \left(-\frac{\alpha w_+}{n} + \alpha w_+ \right)^2 \right) \\ &\qquad - (P_+^t - p_i^t) \cdot \frac{\alpha w_+}{n} - P_-^t \cdot \frac{\alpha w_-}{n} + o(n^{-1}) \Bigg) \\ &= \Lambda_i^t \cdot \left(1 + p_i^t \cdot \left(\alpha w_+ + \left(-\frac{\alpha w_+}{n} + \alpha w_+ \right)^2 \right) - P_+^t \cdot \frac{\alpha w_+}{n} - P_-^t \cdot \frac{\alpha w_-}{n} + o(n^{-1}) \right). \end{aligned}$$

Applying $(-\frac{\alpha w_+}{n} + \alpha w_+)^2 \leqslant (\alpha w_+)^2$ and then replacing P_-^t by $1 - P_+^t$, gives,

$$\begin{aligned} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant \Lambda_i^t \cdot \left(1 + p_i^t \cdot (\alpha w_+ + (\alpha w_+)^2) - P_+^t \cdot \frac{\alpha w_+}{n} - P_-^t \cdot \frac{\alpha w_-}{n} + o(n^{-1}) \right) \\ &= \Lambda_i^t \cdot \left(1 + p_i^t \cdot (\alpha w_+ + (\alpha w_+)^2) + P_+^t \cdot \left(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n} \right) - \frac{\alpha w_-}{n} + o(n^{-1}) \right). \end{aligned} \tag{8.3}$$

By condition $\mathcal{P}_2$, we have $p_i^t \leqslant 1/n$ for any overloaded bin $i \in B_+^t$. Since by assumption, $\delta^t \leqslant 1 - \varepsilon$ we also have that $P_+^t \leqslant (\delta^t \cdot n) \cdot 1/n \leqslant 1 - \varepsilon$. Hence,

$$\begin{aligned} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant \Lambda_i^t \cdot \left(1 + p_i^t \cdot (\alpha w_+ + (\alpha w_+)^2) + (1-\varepsilon) \cdot \left(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n} \right) - \frac{\alpha w_-}{n} + o(n^{-1}) \right) \\ &= \Lambda_i^t \cdot \left(1 - \alpha w_+ \cdot \left(\frac{1}{n} - p_i^t \cdot (1 + \alpha w_+) \right) - \frac{\alpha\varepsilon}{n} \cdot (w_- - w_+) + o(n^{-1}) \right). \end{aligned}$$

Case 1.A [$\mathcal{P}_3$ holds]. Then, as for any overloaded bin $p_i^t \leqslant \frac{1-k_1\varepsilon}{n}$ and $w_- \geqslant w_-$,

$$\begin{aligned} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant \Lambda_i^t \cdot \left(1 - \frac{\alpha w_+}{n} \cdot \Big(1 - (1 - k_1\varepsilon) \cdot (1 + \alpha w_+)\Big) - \frac{\alpha\varepsilon}{n} \cdot (w_- - w_+) + o(n^{-1}) \right) \\ &\leqslant \Lambda_i^t \cdot \left(1 - \frac{\alpha w_+}{n} \cdot \Big(1 - (1 - k_1\varepsilon) \cdot (1 + \alpha w_+)\Big) + o(n^{-1}) \right) \\ &= \Lambda_i^t \cdot \left(1 - \frac{\alpha w_+}{n} \cdot \Big(k_1\varepsilon - \alpha w_+ \cdot (1 - k_1\varepsilon)\Big) + o(n^{-1}) \right) \\ &\overset{(a)}{\leqslant} \Lambda_i^t \cdot \left(1 - \frac{\alpha w_+}{n} \cdot \frac{k_1\varepsilon}{2} + o(n^{-1}) \right) \\ &\overset{(b)}{\leqslant} \Lambda_i^t \cdot \left(1 - \frac{\alpha w_+ k_1\varepsilon}{4n} \right), \end{aligned}$$

where in inequality (a) we used that $\frac{k_1 \varepsilon}{2} \geqslant \alpha w_+(1 - k_1\varepsilon)$ (as implied by $\alpha \leqslant \frac{k_1\varepsilon}{2w_+(1-k_1\varepsilon)}$) and in inequality (b) that $\alpha, w_+, \varepsilon, k_1$ are constants.

Case 1.B [$\mathcal{W}_3$ holds]. Applying the bound $p_i^t \leqslant p_+^t \leqslant 1/n$,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] &\leqslant \Lambda_i^t \cdot \Big(1 - \alpha w_+ \cdot \Big(\frac{1}{n} - \frac{1}{n} \cdot (1 + \alpha w_+)\Big) - \frac{\alpha\varepsilon}{n} \cdot (w_- - w_+) + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 + \frac{(\alpha w_+)^2}{n} - \frac{\alpha\varepsilon}{n} \cdot (w_- - w_+) + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{n} \cdot (\varepsilon \cdot (w_- - w_+) - \alpha w_+^2) + o(n^{-1})\Big) \\
&\overset{(a)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{n} \cdot \frac{\varepsilon}{2} \cdot (w_- - w_+) + o(n^{-1})\Big) \\
&\overset{(b)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha\varepsilon}{4n} \cdot (w_- - w_+)\Big),
\end{aligned}$$

where we used in (a) that $\frac{\varepsilon}{2} \cdot (w_- - w_+) \geqslant \alpha w_+^2$ (as implied by $\alpha \leqslant \frac{\varepsilon(w_- - w_+)}{4w_-^2} \leqslant \frac{\varepsilon(w_- - w_+)}{2w_+^2}$) and in (b) that $\alpha, \varepsilon, (w_- - w_+)$ are constants.

So, in both **Case 1.A** and **Case 1.B**, we conclude that there exists a constant $c_3 > 0$ such that,

$$\sum_{i \in B_{++}^t} \mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \sum_{i \in B_{++}^t} \Lambda_i^t \cdot \Big(1 - \frac{2c_3\alpha}{n}\Big).$$

Case 2 [Robustly Underloaded Bins]. For $i \in B_{--}^t$,

$$\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] = \Lambda_i^t \cdot \Big(p_i^t \cdot e^{\alpha w_-/n - \alpha w_-} + (P_-^t - p_i^t) \cdot e^{\alpha w_-/n} + P_+^t \cdot e^{\alpha w_+/n}\Big).$$

Applying the Taylor estimate $e^z \leqslant 1 + z + z^2$, which holds for any $z \leqslant 1.75$, since $\alpha w_+/n \leqslant 1$ (and $\alpha w_-/n \leqslant 1$),

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Lambda_i^t \cdot \Big(1 + p_i^t \cdot \Big(\frac{\alpha w_-}{n} - \alpha w_- + \Big(\frac{\alpha w_-}{n} - \alpha w_-\Big)^2\Big) \\
+ (P_-^t - p_i^t) \cdot \Big(\frac{\alpha w_-}{n} + \Big(\frac{\alpha w_-}{n}\Big)^2\Big) + P_+^t \cdot \Big(\frac{\alpha w_+}{n} + \Big(\frac{\alpha w_+}{n}\Big)^2\Big)\Big).
\end{aligned}$$

By gathering $o(n^{-1})$ terms and rearranging terms we obtain,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] &\leqslant \Lambda_i^t \cdot \Big(1 + p_i^t \cdot \Big(\frac{\alpha w_-}{n} - \alpha w_- + \Big(\frac{\alpha w_-}{n} - \alpha w_-\Big)^2\Big) \\
&\qquad + (P_-^t - p_i^t) \cdot \frac{\alpha w_-}{n} + P_+^t \cdot \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - p_i^t \cdot \Big(\alpha w_- - \Big(\frac{\alpha w_-}{n} - \alpha w_-\Big)^2\Big) + P_-^t \cdot \frac{\alpha w_-}{n} + P_+^t \cdot \frac{\alpha w_+}{n} + o(n^{-1})\Big).
\end{aligned}$$

Applying $(\frac{\alpha w_-}{n} - \alpha w_-)^2 \leqslant (\alpha w_-)^2$ and then replacing P_+^t by $1 - P_-^t$,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] &\leqslant \Lambda_i^t \cdot \Big(1 - p_i^t \cdot (\alpha w_- - (\alpha w_-)^2) + P_-^t \cdot \frac{\alpha w_-}{n} + P_+^t \cdot \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - p_i^t \cdot (\alpha w_- - (\alpha w_-)^2) + P_-^t \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big).
\end{aligned} \tag{8.4}$$

Case 2.A [$\mathcal{P}_3$ holds]. Using that $p_-^t \geqslant \frac{1+k_2\varepsilon}{n}$, $P_-^t \leqslant 1$, and applying this to (8.4) yields (a) below (since $(\alpha w_-)^2 \leqslant \alpha w_-$), and further rearranging gives,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] &\overset{(a)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{1+k_2\varepsilon}{n} \cdot (\alpha w_- - (\alpha w_-)^2) + 1 \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \alpha w_- \cdot \Big(\frac{1+k_2\varepsilon}{n} \cdot (1 - \alpha w_-) - \frac{1}{n}\Big) + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \Big((1+k_2\varepsilon) \cdot (1 - \alpha w_-) - 1\Big) + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \Big(k_2\varepsilon - \alpha w_- \cdot (1+k_2\varepsilon)\Big) + o(n^{-1})\Big) \\
&\overset{(b)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \frac{k_2\varepsilon}{2} + o(n^{-1})\Big) \\
&\overset{(c)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_- k_2\varepsilon}{4n}\Big),
\end{aligned}$$

where we used in (b) that $\frac{k_2\varepsilon}{2} \geqslant \alpha w_- \cdot (1+k_2\varepsilon)$ (as implied by $\alpha \leqslant \frac{k_2\varepsilon}{2w_-(1+k_2\varepsilon)}$) and in (c) that $\alpha, w_-, k_2, \varepsilon$ are constants.

Case 2.B [$\mathcal{W}_3$ and $P_-^t \leqslant 1 - \frac{\varepsilon}{2}$ holds]. Then since $p_i^t \geqslant 1/n$ and $(\alpha w_-)^2 \leqslant \alpha w_-$, (8.4) implies,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,\right] &\leqslant \Lambda_i^t \cdot \Big(1 - \frac{1}{n} \cdot (\alpha w_- - (\alpha w_-)^2) + (1 - \frac{\varepsilon}{2}) \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 + \frac{(\alpha w_-)^2}{n} - \frac{\varepsilon}{2} \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{n} \cdot \Big(\frac{\varepsilon}{2} \cdot (w_- - w_+) - \alpha w_-^2\Big) + o(n^{-1})\Big) \\
&\overset{(a)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{n} \cdot \frac{\varepsilon}{4} \cdot (w_- - w_+) + o(n^{-1})\Big) \\
&\overset{(b)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha\varepsilon}{8n} \cdot (w_- - w_+)\Big).
\end{aligned}$$

where we used in (a) that $\frac{\varepsilon}{4} \cdot (w_- - w_+) \geqslant \alpha w_-^2$ (as implied by $\alpha \leqslant \frac{\varepsilon(w_- - w_+)}{4w_-^2}$) and in (b) that $\alpha, \varepsilon, (w_- - w_+)$ are constants.

Case 2.C [$\mathcal{W}_3$ and $P_-^t > 1 - \frac{\varepsilon}{2}$ holds]. We will derive a similar inequality by a majorization argument. Using (8.4), for the bins in B_{--}^t,

$$\sum_{i \in B_{--}^t} \mathbf{E}[\, \Lambda_i^{t+1} \mid \mathfrak{F}^t \,] \leqslant \sum_{i \in B_{--}^t} \Lambda_i^t \cdot \Big(1 - p_i^t \cdot (\alpha w_- - (\alpha w_-)^2) + P_-^t \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big).$$

Since Λ_i^t is non-decreasing for $i \in B_{--}^t$ (assuming bins are sorted decreasingly according to their load) and p_i^t is non-decreasing, applying Lemma A.2, we can upper bound the above expression by replacing each p_i^t by the average probability $\overline{p}_i^t$ of robustly underloaded bins. Note that because of monotonicity of the allocation distribution p^t, the average probability over B_{--}^t is at least as large as the average probability over B_-^t, so it satisfies

$$\overline{p}_i^t \geqslant \frac{P_-^t}{|B_-^t|} > \frac{1 - \frac{\varepsilon}{2}}{(1-\varepsilon)n} > \frac{1 + \frac{\varepsilon}{2}}{n},$$

where we have used $P_-^t > 1 - \frac{\varepsilon}{2}$, $|B_-^t| \geqslant (1-\varepsilon) \cdot n$ and Claim A.10. Hence, using that

$(\alpha w_-)^2 \leqslant \alpha w_-$

$$\begin{aligned}
\sum_{i \in B^t_{--}} \mathbf{E}[\Lambda_i^{t+1} \mid \mathfrak{F}^t] &\leqslant \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \overline{p}_i^t \cdot (\alpha w_- - (\alpha w_-)^2) + 1 \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\
&= \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \overline{p}_i^t \cdot (\alpha w_- - (\alpha w_-)^2) + \frac{\alpha w_-}{n} + o(n^{-1})\Big) \\
&= \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \alpha w_- \cdot \Big(\overline{p}_i^t \cdot (1 - \alpha w_-) - \frac{1}{n}\Big) + o(n^{-1})\Big) \\
&\leqslant \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \Big(\Big(1 + \frac{\varepsilon}{2}\Big) \cdot (1 - \alpha w_-) - 1\Big) + o(n^{-1})\Big) \\
&= \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \Big(\frac{\varepsilon}{2} - \alpha w_- \cdot \Big(1 + \frac{\varepsilon}{2}\Big)\Big) + o(n^{-1})\Big) \\
&\overset{(a)}{\leqslant} \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{n} \cdot \frac{\varepsilon}{4} + o(n^{-1})\Big) \\
&\overset{(b)}{\leqslant} \sum_{i \in B^t_{--}} \Lambda_i^t \cdot \Big(1 - \frac{\alpha w_-}{4} \cdot \frac{\varepsilon}{2n}\Big),
\end{aligned}$$

where we used in (a) that $\frac{\varepsilon}{4} \geqslant \alpha w_- \cdot (1 + \frac{\varepsilon}{2})$ (as implied by $\alpha \leqslant \frac{\varepsilon}{2 \cdot w_- \cdot (2+\varepsilon)}$) and in (b) that α, w_-, ε are constants. Hence, in all subcases of **Case 2**, we can find a constant $c_3 > 0$,

$$\sum_{i \in B^t_{--}} \mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t\right] \leqslant \sum_{i \in B^t_{--}} \Lambda^t \cdot \Big(1 - \frac{2c_3\alpha}{n}\Big).$$

Case 3 [Swinging Bins]. For the bins in $B^t_{+/-}$, using Lemma 8.3 with $\kappa_1 := c_3$ gives

$$\sum_{i \in B^t_{+/-}} \mathbf{E}[\Lambda_i^{t+1} \mid \mathfrak{F}^t] \leqslant \sum_{i \in B^t_{+/-}} \Lambda_i^t \cdot \Big(1 - \frac{2c_3\alpha}{n}\Big) + 3(c_3 + w_-) \cdot e^{2w_-}.$$

In conclusion, by choosing a sufficiently small constant $c_3 = c_3(w_+, w_-, k_1, k_2, \varepsilon) > 0$ if $\mathcal{P}_3$ holds and $c_3 = c_3(w_+, w_-, \varepsilon) > 0$ if $\mathcal{W}_3$ holds, which satisfies all the above cases, the claim follows. $\square$

Now, we will prove that the expected increase of Λ is bounded by a factor of $\alpha^2 c_4/2n$ at an arbitrary round.

Lemma 8.5. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. Then for the constant $c_4 := 3w_- \cdot e^{2w_-} > 0$ and any constant α satisfying the preconditions of Lemma 8.4, and any $t \geqslant 0$, we have*

$$\mathbf{E}[\Lambda^{t+1} \mid \mathfrak{F}^t] \leqslant \Lambda^t \cdot \Big(1 + \frac{\alpha^2 c_4}{2n}\Big) + c_4.$$

Proof. **Case 1** [Robustly Overloaded Bins]. Using (8.3) in **Case 1** of Lemma 8.4,

$$\mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t\right] \leqslant \Lambda_i^t \cdot \Big(1 + p_i^t \cdot \Big(\alpha w_+ + (\alpha w_+)^2\Big) + P_+^t \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) - \frac{\alpha w_-}{n} + o(n^{-1})\Big).$$

Using that $p_+^t \leqslant 1/n$ and $P_+^t \leqslant 1$,

$$\begin{aligned}
\mathbf{E}\left[\Lambda_i^{t+1} \mid \mathfrak{F}^t\right] &\leqslant \Lambda_i^t \cdot \Big(1 + \frac{1}{n} \cdot \Big(\alpha w_+ + (\alpha w_+)^2\Big) + \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) - \frac{\alpha w_-}{n} + o(n^{-1})\Big) \\
&= \Lambda_i^t \cdot \Big(1 + \frac{1}{n} \cdot (\alpha w_+)^2 + o(n^{-1})\Big) \leqslant \Lambda_i^t \cdot \Big(1 + \frac{2\alpha^2 w_+^2}{n}\Big),
\end{aligned}$$

where in the last step we used that α and w_+ are constants.

Case 2 [Robustly Underloaded Bins]. Using (8.4) in **Case 2** of Lemma 8.4, we get

$$\mathbf{E}\left[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,\right] \leqslant \Lambda_i^t \cdot \Big(1 + p_i^t \cdot (-\alpha w_- + (\alpha w_-)^2) + P_-^t \cdot \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big)$$

Using that $p_-^t \geqslant \frac{1}{n}$, $\alpha w_- \geqslant (\alpha w_-)^2$ and $P_-^t \leqslant 1$,

$$\begin{aligned}\mathbf{E}\left[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,\right] &\leqslant \Lambda_i^t \cdot \Big(1 + \frac{1}{n} \cdot (-\alpha w_- + (\alpha w_-)^2) + \Big(\frac{\alpha w_-}{n} - \frac{\alpha w_+}{n}\Big) + \frac{\alpha w_+}{n} + o(n^{-1})\Big) \\ &= \Lambda_i^t \cdot \Big(1 + \frac{1}{n} \cdot (\alpha w_-)^2 + o(n^{-1})\Big) \leqslant \Lambda_i^t \cdot \Big(1 + \frac{2\alpha^2 w_-^2}{n}\Big),\end{aligned}$$

where in the last step we used that α and w_- are constants.

Case 3 [Swinging Bins]. Using Lemma 8.3 with $\kappa_1 = 0$ gives

$$\sum_{i \in B_{+/-}^t} \mathbf{E}[\,\Lambda_i^{t+1} \mid \mathfrak{F}^t\,] \leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t + 3w_- \cdot e^{2w_-} \leqslant \sum_{i \in B_{+/-}^t} \Lambda_i^t \cdot \Big(1 + \frac{\alpha^2 w_-^2}{2n}\Big) + 3w_- \cdot e^{2w_-}.$$

Thus, aggregating over the three cases and choosing $c_4 := 3w_- \cdot e^{2w_-} \geqslant 4w_-^2$, gives the result. □

We now combine the statements (and constants) from Lemmas 8.4 and 8.5 into a single corollary.

Corollary 8.6. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process, and let $\varepsilon \in (0, 1)$ be any constant. Choose $c := \max\Big(\frac{3(c_3+w_-)\cdot e^{2w_-}}{\alpha c_3}, \frac{2}{\alpha^2}\Big) > 1$, for the constants $c_3(\varepsilon)$ and $\alpha(\varepsilon)$ as defined in Lemma 8.4. Then for any $t \geqslant 0$,*

$$\mathbf{E}[\,\Lambda^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t, \Lambda^t \geqslant c \cdot n\,] \leqslant \Lambda^t \cdot \Big(1 - \frac{c_3 \alpha}{n}\Big).$$

More generally, for any $t \geqslant 0$, and $c_4 := 3w_- \cdot e^{2w_-} > 0$, we have

$$\mathbf{E}[\,\Lambda^{t+1} \mid \mathfrak{F}^t, \Lambda^t \geqslant c \cdot n\,] \leqslant \Lambda^t \cdot \Big(1 + \frac{\alpha^2 c_4}{n}\Big).$$

Proof. For the first statement, using Lemma 8.4, where $c_3' := 3(c_3 + w_-) \cdot e^{2w_-}$, gives

$$\mathbf{E}[\,\Lambda^{t+1} \mid \mathfrak{F}^t, \mathcal{G}^t, \Lambda^t \geqslant c{\cdot}n\,] \leqslant \Lambda^t{\cdot}\Big(1 - \frac{2c_3\alpha}{n}\Big) + c_3' = \Lambda^t{\cdot}\Big(1 - \frac{c_3\alpha}{n}\Big) + \Big(c_3' - \Lambda^t{\cdot}\frac{c_3\alpha}{n}\Big) \leqslant \Lambda^t{\cdot}\Big(1 - \frac{c_3\alpha}{n}\Big).$$

For the second statement, by applying Lemma 8.5 we have

$$\mathbf{E}[\,\Lambda^{t+1} \mid \mathfrak{F}^t, \Lambda^t \geqslant c \cdot n\,] \leqslant \Lambda^t \cdot \Big(1 + \frac{\alpha^2 c_4}{2n}\Big) + c_4 \leqslant \Lambda^t \cdot \Big(1 + \frac{\alpha^2 c_4}{2n}\Big) + \Lambda^t \cdot \frac{\alpha^2 c_4}{2n} \leqslant \Lambda^t \cdot \Big(1 + \frac{\alpha^2 c_4}{n}\Big),$$

as claimed. □

The next lemma shows that loads cannot deviate too wildly within a $\Theta(n \log n)$-length interval following a round with small exponential potential.

Lemma 8.7. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. For any constant $\kappa > 0$, and any rounds t_0, t_1 such that $t_0 \leqslant t_1 \leqslant t_0 + \kappa \cdot n \log n$,*

$$\mathbf{Pr}\left[\,\max_{t \in [t_0, t_1]} \max_{i \in [n]} |y_i^t| \leqslant \log^2 n \,\middle|\, \mathfrak{F}^{t_0},\ \Lambda^{t_0} \leqslant n^2\,\right] \geqslant 1 - n^{-12}.$$

Proof. Consider the sequence $(\mathbf{E}[\Lambda^t \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant n^2])_{t=t_0}^{t_1}$, then for every $t \in (t_0, t_1]$, using Lemma 8.5,

$$\mathbf{E}[\Lambda^{t+1} \mid \mathfrak{F}^t] \leqslant \Lambda^t \cdot \Big(1 + \frac{c_4 \alpha^2}{2n}\Big) + c_4.$$

Using the tower law of expectation, we have that

$$\mathbf{E}[\Lambda^{t+1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant n^2] \leqslant \mathbf{E}\left[\Lambda^t \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant n^2\right] \cdot \Big(1 + \frac{c_4 \alpha^2}{2n}\Big) + c_4.$$

Hence, applying Lemma A.12 with $a := 1 + \frac{c_4 \alpha^2}{2n} > 1$ and $b := c_4 > 0$, we get that

$$\begin{aligned}
\mathbf{E}[\Lambda^t \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant n^2] &\leqslant \Lambda^{t_0} \cdot a^{t-t_0} + b \cdot \sum_{s=t_0}^{t-1} a^{s-t_0} \\
&\leqslant n^2 \cdot \Big(1 + \frac{\alpha^2 c_4}{n}\Big)^{t_1 - t_0} + c_4 \cdot (t_1 - t_0) \cdot \Big(1 + \frac{\alpha^2 c_4}{n}\Big)^{t_1 - t_0} \\
&\overset{(a)}{\leqslant} n^2 \cdot n^{\kappa \cdot \alpha^2 \cdot c_4} + c_4 \cdot (\kappa n \log n) \cdot n^{\kappa \cdot \alpha^2 \cdot c_4} \leqslant n^{3 + \kappa \cdot \alpha^2 \cdot c_4}.
\end{aligned}$$

where in (a) we used that $1 + z \leqslant e^z$ for any z.

Using Markov's inequality, $\mathbf{Pr}\left[\Lambda^t \leqslant n^{3+\kappa \cdot \alpha^2 \cdot c_4 + 14} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant n^2\right] \geqslant 1 - n^{-14}$ for any $t \in [t_0, t_1]$, which implies

$$\mathbf{Pr}\left[\max_{i \in [n]} |y_i^t| \leqslant \frac{1}{\alpha} \cdot (\kappa \cdot \alpha^2 \cdot c_4 + 17) \cdot \log n \;\middle|\; \mathfrak{F}^{t_0},\ \Lambda^{t_0} \leqslant n^2\right] \geqslant 1 - n^{-14}.$$

Since $\frac{1}{\alpha} \cdot (\kappa \cdot \alpha^2 \cdot c_4 + 17) \cdot \log n < \log^2 n$ for sufficiently large n, by taking a union bound over all rounds $t \in [t_0, t_1]$ we get the claim. □

8.3 Weak Exponential Potential V

For the massively-loaded case, we will use the exponential potential V with a parameter $\widetilde{\alpha} = \Theta(1/n)$ (to be fixed below) and show that its expectation is $\mathcal{O}(n^4)$. In Lemma 8.9, using Markov's inequality, we will deduce that w.h.p. the gap at an arbitrary round is $\mathcal{O}(n \log n)$.

As in Section 8.2, the analysis will be done over the three possible bins: robustly underloaded bins B_{++}^t with load $y_i^t \geqslant \frac{w_-}{n}$, robustly underloaded bins B_{--}^t with load $y_i^t \leqslant -w_-$ and swinging bins $B_{+/-}^t$ with load $y_i^t \in (-w_-, \frac{w_-}{n})$.

Lemma 8.8. *For any $\mathcal{W}_3$-process set*

$$\widetilde{\alpha} := \min\left\{\frac{k_1}{2 \cdot w_- \cdot (n - k_1)}, \frac{k_2}{2 \cdot w_- \cdot (n + k_2)}\right\},$$

and for any $\mathcal{P}_3$-process set

$$\widetilde{\alpha} := \min\left\{\frac{w_- - w_+}{4 \cdot w_-^2 \cdot n}, \frac{1}{w_- \cdot (4n + 2)}\right\}.$$

Then there exists a constant $c_5 > 0$ such that for any $t \geqslant 0$,

$$\mathbf{E}\left[V^{t+1} \mid \mathfrak{F}^t\right] \leqslant V^t \cdot \Big(1 - \frac{c_5}{n^3}\Big) + 2n.$$

Proof. Following the text before the lemma, we divide $\mathbf{E}\left[V^{t+1} \mid \mathfrak{F}^t\right]$ into three cases of bins as follows,

$$\mathbf{E}\left[V^{t+1} \mid \mathfrak{F}^t\right] = \sum_{i \in B_{++}^t} \mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t\right] + \sum_{i \in B_{--}^t} \mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t\right] + \sum_{i \in B_{+/-}^t} \mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t\right], \tag{8.5}$$

and upper bound each of the three sums separately. Note that if $\delta^t = 1$, then all bins have $y^t = 0$, so all bins belong to **Case 3**. So, in **Case 1** and **Case 2**, we can assume that $1/n \leqslant \delta^t \leqslant 1 - 1/n$.
Case 1 [Robustly Overloaded Bins]. For $i \in B_{++}^t$,

$$\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] = V_i^t \cdot \Big(p_i^t \cdot e^{-\widetilde{\alpha} w_+/n + \widetilde{\alpha} w_+} + (P_+^t - p_i^t) \cdot e^{-\widetilde{\alpha} w_+/n} + P_-^t \cdot e^{-\widetilde{\alpha} w_-/n} \Big)$$

Applying the Taylor estimate $e^z \leqslant 1 + z + z^2$, which holds for any $z \leqslant 1.75$, since $\widetilde{\alpha} \leqslant \frac{1}{n}$,

$$\begin{aligned}\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] \leqslant V_i^t \cdot \Big(1 + p_i^t \cdot \Big(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+ + \Big(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+ \Big)^2 \Big) \\ + (P_+^t - p_i^t) \cdot \Big(-\frac{\widetilde{\alpha} w_+}{n} + \Big(\frac{\widetilde{\alpha} w_+}{n} \Big)^2 \Big) + P_-^t \cdot \Big(-\frac{\widetilde{\alpha} w_-}{n} + \Big(\frac{\widetilde{\alpha} w_-}{n} \Big)^2 \Big) \Big).\end{aligned}$$

Using the fact that $\frac{\widetilde{\alpha}^2}{n^2} = \mathcal{O}(n^{-4})$ and then subsequently rearranging terms we obtain

$$\begin{aligned}\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant V_i^t \cdot \Big(1 + p_i^t \cdot \Big(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+ + \Big(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+ \Big)^2 \Big) \\ &\qquad - (P_+^t - p_i^t) \cdot \frac{\widetilde{\alpha} w_+}{n} - P_-^t \cdot \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3}) \Big) \\ &= V_i^t \cdot \Big(1 + p_i^t \cdot \Big(\widetilde{\alpha} w_+ + \Big(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+ \Big)^2 \Big) - P_+^t \cdot \frac{\widetilde{\alpha} w_+}{n} - P_-^t \cdot \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3}) \Big).\end{aligned}$$

Applying $(-\frac{\widetilde{\alpha} w_+}{n} + \widetilde{\alpha} w_+)^2 \leqslant (\widetilde{\alpha} w_+)^2$ then gives

$$\begin{aligned}\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant V_i^t \cdot \Big(1 + p_i^t \cdot (\widetilde{\alpha} w_+ + (\widetilde{\alpha} w_+)^2) - P_+^t \cdot \frac{\widetilde{\alpha} w_+}{n} - P_-^t \cdot \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3}) \Big) \\ &= V_i^t \cdot \Big(1 + p_i^t \cdot (\widetilde{\alpha} w_+ + (\widetilde{\alpha} w_+)^2) + P_+^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n} \Big) - \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3}) \Big),\end{aligned}$$

where in the first equality we replaced P_-^t by $1 - P_+^t$ and rearranged. Using $P_+^t \leqslant \delta^t \leqslant 1 - \frac{1}{n}$,

$$\begin{aligned}\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant V_i^t \cdot \Big(1 + p_i^t \cdot (\widetilde{\alpha} w_+ + (\widetilde{\alpha} w_+)^2) + \Big(1 - \frac{1}{n} \Big) \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n} \Big) - \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3}) \Big) \\ &= V_i^t \cdot \Big(1 - \widetilde{\alpha} w_+ \cdot \Big(\frac{1}{n} - p_i^t \cdot (1 + \widetilde{\alpha} w_+) \Big) - \frac{\widetilde{\alpha}}{n^2} \cdot (w_- - w_+) + o(n^{-3}) \Big). \qquad (8.6)\end{aligned}$$

Case 1.A [$\mathcal{P}_3$ holds]. The bounds $p_+^t \leqslant \frac{1}{n} - \frac{k_1 \cdot (1-\delta^t)}{n} \leqslant \frac{1}{n} - \frac{k_1}{n^2}$ and $w_- \geqslant w_+$ applied to (8.6) above give

$$\begin{aligned}\mathbf{E}\left[V_i^{t+1} \mid \mathfrak{F}^t \right] &\leqslant V_i^t \cdot \Big(1 - \widetilde{\alpha} w_+ \cdot \Big(\frac{1}{n} - \Big(\frac{1}{n} - \frac{k_1}{n^2} \Big) \cdot (1 + \widetilde{\alpha} w_+) \Big) + o(n^{-3}) \Big) \\ &= V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_+}{n} \cdot \Big(\frac{k_1}{n} - \Big(1 - \frac{k_1}{n} \Big) \cdot (\widetilde{\alpha} w_+) \Big) + o(n^{-3}) \Big) \\ &\overset{(a)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_+}{n} \cdot \frac{k_1}{2n} + o(n^{-3}) \Big) \\ &\overset{(b)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_+ k_1}{4n^2} \Big),\end{aligned}$$

where we used in (a) that $\frac{k_1}{2n} \leqslant (1 - \frac{k_1}{n}) \cdot \widetilde{\alpha} w_+$ (as implied by $\widetilde{\alpha} \leqslant \frac{k_1}{2 \cdot w_- \cdot (n - k_1)}$) and in (b) that $\widetilde{\alpha} = \Omega(n^{-1})$.

Case 1.B [$\mathcal{W}_3$ holds]. Applying the bound $p_+^t \leqslant \frac{1}{n}$ to (8.6) yields

$$\begin{aligned}
\mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] &\leqslant V_i^t \cdot \Big(1 - \widetilde{\alpha} w_+ \cdot \Big(\frac{1}{n} - \frac{1}{n} \cdot (1 + \widetilde{\alpha} w_+)\Big) - \frac{\widetilde{\alpha}}{n^2} \cdot (w_- - w_+) + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 + \frac{(\widetilde{\alpha} w_+)^2}{n} - \frac{\widetilde{\alpha}}{n^2} \cdot (w_- - w_+) + o(n^{-3})\Big) \\
&\overset{(a)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha}}{2n^2} \cdot (w_- - w_+) + o(n^{-3})\Big) \\
&\overset{(b)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha}}{4n^2} \cdot (w_- - w_+)\Big),
\end{aligned}$$

where we used in (a) that $\frac{(\widetilde{\alpha} w_+)^2}{2n} \leqslant \frac{\widetilde{\alpha}}{n^2} \cdot (w_- - w_+)$ (as implied by $\widetilde{\alpha} \leqslant \frac{w_- - w_+}{4 \cdot w_-^2 \cdot n} \leqslant \frac{w_- - w_+}{2 \cdot w_+^2 \cdot n}$) and in (b) that $\widetilde{\alpha} = \Omega(n^{-1})$.

So, in both **Case 1.A** and **Case 1.B**, we get for some constant $c_5 > 0$,

$$\sum_{i \in B_{++}^t} \mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \sum_{i \in B_{++}^t} V_i^t \cdot \Big(1 - \frac{c_5}{n^3}\Big).$$

Case 2 [Robustly Underloaded Bins]. For any $i \in B_{--}^t$, using again the Taylor estimate $e^z \leqslant 1 + z + z^2$ (since $\widetilde{\alpha} = \Theta(1/n)$) yields

$$\begin{aligned}
\mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] &= V_i^t \cdot (p_i^t \cdot e^{\widetilde{\alpha} w_-/n - \widetilde{\alpha} w_-} + (P_-^t - p_i^t) \cdot e^{\widetilde{\alpha} w_-/n} + P_+^t \cdot e^{\widetilde{\alpha} w_+/n}) \\
&\leqslant V_i^t \cdot \Big(1 + p_i^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \widetilde{\alpha} w_- + \Big(\frac{\widetilde{\alpha} w_-}{n} - \widetilde{\alpha} w_-\Big)^2\Big) \\
&\qquad + (P_-^t - p_i^t) \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} + \Big(\frac{\widetilde{\alpha} w_-}{n}\Big)^2\Big) + P_+^t \cdot \Big(\frac{\widetilde{\alpha} w_+}{n} + \Big(\frac{\widetilde{\alpha} w_+}{n}\Big)^2\Big)\Big)
\end{aligned}$$

Using the fact that $\frac{\widetilde{\alpha}^2}{n^2} = \mathcal{O}(n^{-4})$ and then substituting $P_+^t = 1 - P_-^t$ yields

$$\begin{aligned}
&\leqslant V_i^t \cdot \Big(1 + p_i^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \widetilde{\alpha} w_- + \Big(\frac{\widetilde{\alpha} w_-}{n} - \widetilde{\alpha} w_-\Big)^2\Big) \\
&\qquad + (P_-^t - p_i^t) \cdot \frac{\widetilde{\alpha} w_-}{n} + P_+^t \cdot \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 - p_i^t \cdot \Big(\widetilde{\alpha} w_- - \Big(-\frac{\widetilde{\alpha} w_-}{n} + \widetilde{\alpha} w_-\Big)^2\Big) + P_-^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big).
\end{aligned}$$

Now, $(-\frac{\widetilde{\alpha} w_-}{n} + \widetilde{\alpha} w_-)^2 \leqslant (\widetilde{\alpha} w_-)^2$ implies that

$$\mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant V_i^t \cdot \Big(1 - p_i^t \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + P_-^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big). \tag{8.7}$$

Case 2.A [$\mathcal{P}_3$ holds]. Using that $p_i^t \geqslant \frac{1 + k_2 \cdot \delta^t}{n} \geqslant \frac{1 + k_2 \cdot \frac{1}{n}}{n} \geqslant \frac{1}{n} + \frac{k_2}{n^2}$, $P_-^t \leqslant 1$, and applying this to the previous inequality yields (a) below (since $(\widetilde{\alpha} w_-)^2 \leqslant \widetilde{\alpha} w_-$ the factor after p_i^t is positive), and further rearranging gives

$$\begin{aligned}
\mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] &\overset{(a)}{\leqslant} V_i^t \cdot \Big(1 - \Big(\frac{1}{n} + \frac{k_2}{n^2}\Big) \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + 1 \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 - \widetilde{\alpha} w_- \cdot \Big(\Big(\frac{1}{n} + \frac{k_2}{n^2}\Big) \cdot (1 - \widetilde{\alpha} w_-) - \frac{1}{n}\Big) + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_-}{n} \cdot \Big(\frac{k_2}{n} - \Big(1 + \frac{k_2}{n}\Big) \cdot \widetilde{\alpha} w_-\Big) + o(n^{-3})\Big) \\
&\overset{(b)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_-}{n} \cdot \frac{k_2}{2n} + o(n^{-3})\Big) \\
&\overset{(c)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_- k_2}{4n^2}\Big).
\end{aligned}$$

where we used in (b) that $\frac{k_2}{2n} \geqslant (1+\frac{k_2}{n}) \cdot \widetilde{\alpha} w_-$ (equivalent to $\widetilde{\alpha} \leqslant \frac{k_2}{2 \cdot w_- \cdot (n+k_2)}$) and in (c) that $\alpha = \Omega(n^{-1})$.

Case 2.B [$\mathcal{W}_3$ holds and $P_-^t \leqslant 1 - \frac{1}{2n}$]. Applying $p_i^t \geqslant 1/n$ and $P_-^t \leqslant 1 - \frac{1}{2n}$ to (8.7) yields (a) below, and then further rearranging gives

$$\begin{aligned}
\mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] &\overset{(a)}{\leqslant} V_i^t \cdot \Big(1 - \frac{1}{n} \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + \Big(1 - \frac{1}{2n}\Big) \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 - \Big(\frac{1}{n} \cdot (1 - \widetilde{\alpha} w_-) - \frac{1}{n}\Big) \cdot \widetilde{\alpha} w_- - \frac{1}{2n} \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + o(n^{-3})\Big) \\
&= V_i^t \cdot \Big(1 + \frac{(\widetilde{\alpha} w_-)^2}{n} - \frac{\widetilde{\alpha}}{2n^2} \cdot (w_- - w_+) + o(n^{-3})\Big) \\
&\overset{(b)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha}}{4n^2} \cdot (w_- - w_+) + o(n^{-3})\Big) \\
&\overset{(c)}{\leqslant} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha}}{8n^2} \cdot (w_- - w_+)\Big),
\end{aligned}$$

where we used in (b) that $\frac{\widetilde{\alpha}}{4n^2} \cdot (w_- - w_+) \geqslant \frac{(\widetilde{\alpha} w_-)^2}{n}$ (equivalent to $\widetilde{\alpha} \leqslant \frac{w_- - w_+}{4 \cdot w_-^2 \cdot n}$) and in (c) that $\widetilde{\alpha} = \Omega(n^{-1})$ and $w_- - w_+$ is a positive constant.

Case 2.C [$\mathcal{W}_3$ holds and $P_-^t > 1 - \frac{1}{2n}$]. We will derive a similar inequality as in the previous cases, but we will take the sum over all $i \in B_{--}^t$ first to apply a majorization argument. Aggregating the contributions over all bins in B_{--}^t yields (taking the sum of (8.7)),

$$\begin{aligned}
\sum_{i \in B_{--}^t} \mathbf{E}[\, V_i^{t+1} \mid \mathfrak{F}^t \,] &\leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - p_i^t \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + P_-^t \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big) \\
&\leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - p_i^t \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + 1 \cdot \Big(\frac{\widetilde{\alpha} w_-}{n} - \frac{\widetilde{\alpha} w_+}{n}\Big) + \frac{\widetilde{\alpha} w_+}{n} + o(n^{-3})\Big) \\
&= \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - p_i^t \cdot (\widetilde{\alpha} w_- - (\widetilde{\alpha} w_-)^2) + \frac{\widetilde{\alpha} w_-}{n} + o(n^{-3})\Big) \\
&= \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - \widetilde{\alpha} w_- \cdot \Big(p_i^t \cdot (1 - \widetilde{\alpha} w_-) - \frac{1}{n}\Big) + o(n^{-3})\Big),
\end{aligned}$$

Since V_i^t is non-decreasing for $i \in B_{--}^t$ (recall the bins are labeled in non-increasing order) and p_i^t is non-decreasing, applying Lemma A.2, we can upper bound the above expression by replacing each p_i^t by the average probability $\overline{p}_i^t$ of robustly underloaded bins. Note that because of monotonicity of p_i^t, the average probability over B_{--}^t is at least as large as the average probability over B_-^t, so it satisfies

$$\overline{p}_i^t \geqslant \frac{P_-^t}{|B_-^t|} \geqslant \frac{1 - \frac{1}{2n}}{n-1} = \frac{1 - \frac{1}{2n}}{1 - \frac{1}{n}} \cdot \frac{1}{n} \geqslant \frac{1}{n} + \frac{1}{2n^2},$$

where we have used the case assumption $P_-^t > 1 - \frac{1}{2n}$ and the invariant $|B_-^t| \leqslant n-1$ and Claim A.10 for $\varepsilon := \frac{1}{n}$ (and $n > 1$). Hence, since the factor after p_i^t is positive as $\widetilde{\alpha} w_- \leqslant 1$, we get

$$\begin{aligned}
\sum_{i \in B_{--}^t} \mathbf{E}[\, V_i^{t+1} \mid \mathfrak{F}^t \,] &\leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - \widetilde{\alpha} w_- \cdot \Big(\Big(\frac{1}{n} + \frac{1}{2n^2}\Big) \cdot (1 - \widetilde{\alpha} w_-) - \frac{1}{n}\Big) + o(n^{-3})\Big) \\
&= \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_-}{n} \cdot \Big(\frac{1}{2n} - \widetilde{\alpha} w_- \cdot \Big(1 + \frac{1}{2n}\Big)\Big) + o(n^{-3})\Big).
\end{aligned}$$

Now, since $\frac{1}{4n} \geqslant \widetilde{\alpha} w_- \cdot (1 + \frac{1}{2n})$ (as implied by $\widetilde{\alpha} \leqslant \frac{1}{w_- \cdot (4n+2)}$) we have

$$\sum_{i \in B_{--}^t} \mathbf{E}[\, V_i^{t+1} \mid \mathfrak{F}^t \,] \leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - \frac{\widetilde{\alpha} w_-}{n} \cdot \frac{1}{4n} + o(n^{-3})\Big).$$

Finally as $\widetilde{\alpha} = \Omega(n^{-1})$ for the contribution of the robustly underloaded bins we have

$$\sum_{i \in B_{--}^t} \mathbf{E}[\, V_i^{t+1} \mid \mathfrak{F}^t \,] \leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \left(1 - \frac{\widetilde{\alpha} w_-}{16 n^2}\right).$$

So, in all **Case 2.A**, **Case 2.B** and **Case 2.C**, we get for some constant $c_5 > 0$,

$$\sum_{i \in B_{--}^t} \mathbf{E}\left[\, V_i^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \sum_{i \in B_{--}^t} V_i^t \cdot \Big(1 - \frac{c_5}{n^3}\Big).$$

Case 3 [Swinging Bins]. For bins $B_{+/-}^t$ with load $y_i^t \in (-w_-, \frac{w_-}{n})$, we have,

$$\sum_{i \in B_{+/-}^t} \mathbf{E}[\, V_i^{t+1} \mid \mathfrak{F}^t \,] \leqslant \sum_{i \in B_{+/-}^t} e^{2\widetilde{\alpha} w_-} \leqslant 2n \leqslant 2n + \sum_{i \in B_{+/-}^t} V_i^t \cdot \Big(1 - \frac{c_5}{n^3}\Big).$$

Applying the above inequalities for the six cases to (8.5), we get the claim of the lemma. □

Lemma 8.9. *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process, it holds that for any $t \geqslant 0$,*

$$\mathbf{E}\left[\, V^t \,\right] \leqslant \frac{2}{c_5} \cdot n^4.$$

Proof of Lemma 8.9. We will prove this by induction. The base case follows trivially, since $V^0 = n$. Assume that $\mathbf{E}\left[\, V^t \,\right] \leqslant n^4 \cdot \frac{2}{c_5}$ for some $t \geqslant 0$, then by Lemma 8.8 and the tower property of expectation,

$$\mathbf{E}\left[\, V^{t+1} \,\right] \leqslant \mathbf{E}\left[\, V^t \,\right] \cdot \left(1 - \frac{c_5}{n^3}\right) + 2n \leqslant \frac{2}{c_5} \cdot n^4 - \frac{2n^4}{c_5} \cdot \frac{c_5}{n^3} + 2n = \frac{2}{c_5} \cdot n^4. \qquad \square$$

We can now apply the above results to deduce that there is a poly(n) gap at an arbitrary round. This will be used in the proof of Theorem 4.15 as the starting point of the "recovery phase" (Lemma 9.5).

Lemma 8.10. *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process, there is some constant $c_6 > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant c_6 \cdot n \log n \,\right] \geqslant 1 - n^{-12},$$

and so, for any constant $\alpha < 1$,

$$\mathbf{Pr}\left[\, \Lambda^m \leqslant \exp(2 c_6 n \log n) \,\right] \geqslant 1 - n^{-12}.$$

Proof. Using Lemma 8.9, we have that

$$\mathbf{E}\left[\, V^m \,\right] \leqslant \frac{2}{c_5} \cdot n^4.$$

By Markov's inequality, we have $\mathbf{Pr}\left[\, V^m \leqslant \frac{2}{c_5} \cdot n^{16} \,\right] \geqslant 1 - n^{-12}$. By taking the log of the potential function we obtain

$$\mathbf{Pr}\left[\, \max_{i \in [n]} |y_i^m| \leqslant \log\left(\frac{2}{c_5}\right) + \frac{16}{\widetilde{\alpha}} \cdot \log n \,\right] \geqslant 1 - n^{-12},$$

thus, as $\widetilde{\alpha} = \Theta(1/n)$, we have $\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant c_6 n \log n \,\right] \geqslant 1 - n^{-12}$ for some constant $c_6 > 0$. Finally, since $\alpha < 1$, when $\max_{i \in [n]} |y_i^m| \leqslant c_6 n \log n$, we get $\Lambda^m \leqslant n \cdot \exp(c_6 n \log n) \leqslant \exp(2 c_6 n \log n)$ and deduce the second statement. □

9 Analysis of Non-Filling Processes

In this section, we complete the proof that any $\mathcal{P}_3 \cap \mathcal{W}_2$ or $\mathcal{P}_2 \cap \mathcal{W}_3$-process satisfies $\mathrm{Gap}(m) = \mathcal{O}(\log n)$ w.h.p. at an arbitrary round m. As outlined in Section 6, the proof consists of a recovery phase and a stabilization phase. Fig. 4 depicts these two phases.

In the recovery phase (Section 9.3), we prove that starting at $t_0 = m - \Theta(n^3 \log^4 n)$ with $V^{t_0} = \mathrm{poly}(n)$ (i.e., $\Lambda^{t_0} \leqslant \exp(2c_6 n \log n)$), w.h.p. there exists a round $s \in [t_0, m]$ with $\Lambda^s < cn$. We do this by proving that with constant probability if $\Lambda^{t_0} \leqslant \exp(2c_6 n \log n)$, then there is a constant fraction of rounds $r \in [t_0, t_0 + n^3 \log^3 n]$ with $\delta^r \in (\varepsilon, 1-\varepsilon)$. By analyzing an "adjusted version" $\widetilde{\Lambda}$ of the exponential potential Λ, taking advantage of the fact that Λ^t decreases in expectation when $\delta^t \in (\varepsilon, 1-\varepsilon)$ and increases at most by a smaller factor otherwise (Corollary 8.6), we show that there exists $s \in [t_0, t_0 + n^3 \log^3 n]$ such that $\Lambda^s < cn$. By repeating this argument $\Theta(\log n)$ times we amplify the probability to get that w.h.p. there exists $s \in [t_0, m]$ with $\Lambda^s < cn$.

Then, in the stabilization phase (Section 9.4), we first show that if $\Lambda^s < 2cn$ for any round s, then $\Lambda^t < cn$ for some round $t \in (s, s+\Theta(n \log n)]$. We do this by proving that w.h.p. if $\Lambda^s < cn$, then there is a constant fraction of rounds $r \in [s, s + \Theta(n \log n)]$ with $\delta^r \in (\varepsilon, 1-\varepsilon)$. Again, by analyzing an "adjusted version" $\widetilde{\Lambda}$ of the exponential potential Λ, we show that w.h.p. there exists $t \in (s, s+\Theta(n \log n)]$ such that $\Lambda^t < cn$. Next, we take the union bound over the remaining $\mathcal{O}(n^3 \log^4 n)$ rounds, which gives $\Lambda^{r_1} \leqslant cn$ w.h.p. at some $r_1 \in [m, m+\Theta(n \log n)]$ and $\Lambda^{r_2} \leqslant cn$ w.h.p. at some $r_2 \in [m - \Theta(n \log n), m]$, this in turn implies that $\max_{i \in [n]} |y_i^m| = \mathcal{O}(\log n)$.

In order to complete the analysis in Sections 9.3 and 9.4 described above, we must first give some definitions and establish some technical tools. In particular, in Section 9.1 we shall prove that the new adjusted exponential potential function $\widetilde{\Lambda}^t$ is a super-martingale. In Section 9.2 we then prove bounds on the random variables involved in this new potential so that we can utilize the super-martingale property to establish a drop in potential later in the proof.

In the following, we consider an arbitrary round $t_0 \geqslant 0$ which will be starting point of our analysis. For integers $s \geqslant t_0 \geqslant 0$ and some arbitrary $\varepsilon \in (0,1)$, to be fixed later in Lemma 9.7, we let $c(\varepsilon) > 1$ be the constant from Corollary 8.6, and with it we define the event

$$\mathcal{E}_{t_0}^s := \bigcap_{r \in [t_0, s]} \{\Lambda^r \geqslant cn\}.$$

Showing that this event does not hold for suitable t_0, s is a large part of proving the gap bound. For rounds $s \geqslant t_0$, we let $G_{t_0}^s := G_{t_0}^s(\varepsilon)$ be the number of rounds $r \in [t_0, s]$ with $\delta^r \in (\varepsilon, 1-\varepsilon)$, and similarly we define $B_{t_0}^s := (s - t_0 + 1) - G_{t_0}^s$.

We now introduce the *adjusted exponential potential function* $\widetilde{\Lambda}_{t_0}^s$ which involves the random variables and event above. Let $c_3(\varepsilon)$ be the constant in Lemma 8.4 and $0 < \gamma \leqslant 1$ be an arbitrary constant (fixed in Lemma 9.5). Then we set $\widetilde{\Lambda}_{t_0}^{t_0} := \Lambda^{t_0}$ and, for any $s > t_0$, we define the sequence

$$\widetilde{\Lambda}_{t_0}^s := \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{c_3 \alpha \gamma}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{c_3 \alpha}{n} \cdot G_{t_0}^{s-1}\right). \tag{9.1}$$

In this section we shall fix many of the constants such as α. We begin by setting

$$C := \left\lceil \frac{8c_2}{c_1} \right\rceil + 1,$$

where c_1, c_2 are the constants from Lemma 6.2.

We also define $\widetilde{G}_{t_0}^s := \widetilde{G}_{t_0}^s(C)$ to be the number of rounds $r \in [t_0, s]$ with $\Delta^r \leqslant C \cdot n$.

9.1 The Adjusted Exponential Potential is a Super-Martingale

We now show that, for a suitably choice of parameters, the sequence defined by (9.1) forms a super-martingale.

Lemma 9.1. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. Let $\varepsilon \in (0,1)$ and $0 < \gamma \leqslant 1$ be arbitrary constants, and let $c_3 := c_3(\varepsilon)$ be the constant in Lemma 8.4. Further, let $0 < \alpha \leqslant c_3\gamma/(3w_- \cdot e^{2w_-})$ be a constant which additionally meets the conditions of Lemma 8.4. Then for any $t_0 \geqslant 0$, we have for any $s \geqslant t_0$ that*

$$\mathbf{E}[\,\widetilde{\Lambda}_{t_0}^{s+1} \mid \mathfrak{F}^s\,] \leqslant \widetilde{\Lambda}_{t_0}^{s}.$$

Proof. We see that $\mathbf{E}[\,\widetilde{\Lambda}_{t_0}^{s+1} \mid \mathfrak{F}^s\,]$ is given by

$$\begin{aligned}
&\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s\,] \cdot \exp\Big(-\frac{c_3\alpha\gamma}{n} \cdot B_{t_0}^s\Big) \cdot \exp\Big(\frac{c_3\alpha}{n} \cdot G_{t_0}^s\Big) \\
&= \mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s\,] \cdot \exp\left(\frac{c_3\alpha\gamma}{n} \cdot \Big(\frac{\gamma+1}{\gamma} \cdot \mathbf{1}_{\mathcal{G}^s} - 1\Big)\right) \cdot \exp\Big(-\frac{c_3\alpha\gamma}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(\frac{c_3\alpha}{n} \cdot G_{t_0}^{s-1}\Big).
\end{aligned}$$

Thus, we see that it suffices to prove that

$$\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s\,] \cdot \exp\left(\frac{c_3\alpha\gamma}{n} \cdot \Big(\frac{\gamma+1}{\gamma} \cdot \mathbf{1}_{\mathcal{G}^s} - 1\Big)\right) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}. \tag{9.2}$$

To show (9.2), we consider two cases based on whether $\mathcal{G}^s$ holds.

Case 1 [$\mathcal{G}^s$ holds]. Recall that the event $\mathcal{G}^s$ means $\delta^s \in (\varepsilon, 1-\varepsilon)$ holds. Further, we are additionally conditioning on the event $\mathcal{E}_{t_0}^s$, via the indicator, and so $\Lambda^t > cn$ holds for any round $t \in [t_0, s]$. Thus we can use the upper bound from Corollary 8.6 to give

$$\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s, \mathcal{G}^s\,] \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \Big(1 - \frac{c_3\alpha}{n}\Big) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\Big(-\frac{c_3\alpha}{n}\Big).$$

Hence, since in this case $\mathbf{1}_{\mathcal{G}^s} = 1$, the left hand side of (9.2) is equal to

$$\begin{aligned}
\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s, \mathcal{G}^s\,] \cdot \exp\left(\frac{c_3\alpha\gamma}{n} \cdot \Big(\frac{\gamma+1}{\gamma} - 1\Big)\right) &\leqslant \Big(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\Big(-\frac{c_3\alpha}{n}\Big)\Big) \cdot \exp\Big(\frac{c_3\alpha}{n}\Big) \\
&= \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.
\end{aligned}$$

Case 2 [$\mathcal{G}^s$ does not hold]. Recall from Corollary 8.6 that $c_4 = 3w_- \cdot e^{2w_-}$, thus our condition on α can be expressed as $\alpha \leqslant \frac{c_3\gamma}{c_4}$. The second inequality of Corollary 8.6 then implies

$$\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s, \neg\mathcal{G}^s\,] \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \left(1 + \frac{c_4\alpha^2}{n}\right) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \Big(1 + \frac{c_3\alpha\gamma}{n}\Big) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\Big(\frac{c_3\alpha\gamma}{n}\Big).$$

Hence, since in this case $\mathbf{1}_{\mathcal{G}^s} = 0$, the left hand side of (9.2) is equal to

$$\begin{aligned}
\mathbf{E}[\,\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathfrak{F}^s, \neg\mathcal{G}^s\,] \cdot \exp\Big(\frac{c_3\alpha\gamma}{n} \cdot (-1)\Big) &\leqslant \Big(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\Big(\frac{c_3\alpha\gamma}{n}\Big)\Big) \cdot \exp\Big(-\frac{c_3\alpha\gamma}{n}\Big) \\
&= \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.
\end{aligned}$$

Since (9.2) holds in either case, we deduce that $(\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0}$ forms a super-martingale. □

9.2 Taming the Mean Quantile and Absolute Value Potential

We have seen in the previous section that if we augment the exponential potential Λ^t with some terms involving the random variable $G_{t_0}^s$, then the resulting potential $\widetilde{\Lambda}^t$ is a super-martingale. Thus it will be useful to control $G_{t_0}^s := G_{t_0}^s(\varepsilon)$, which we recall is the number of rounds $r \in [t_0, s]$ with $\delta^r \in (\varepsilon, 1-\varepsilon)$. This is done in part by controlling $\widetilde{G}_{t_0}^s := \widetilde{G}_{t_0}^s(C)$, which we recall is the number of rounds $r \in [t_0, s]$ with $\Delta^r \leqslant C \cdot n$. In particular, our first lemma shows that for an interval with length at least $5Cn$, if we have a constant fraction of steps t with $\Delta^t \leqslant Cn$, then w.h.p. we also have a constant fraction of steps with $\delta^t \in (\varepsilon, 1-\varepsilon)$.

Lemma 9.2. *Consider any integer constant $C \geqslant 1$ and any two rounds t_0 and t_1 such that $t_0 + 5Cn \leqslant t_1$. Then there exists a constant $\varepsilon := \varepsilon(C) > 0$ such that*

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > \frac{\varepsilon}{5C} \cdot \widetilde{G}_{t_0}^{t_1 - 5Cn} \;\middle|\; \mathfrak{F}^{t_0} \right] \geqslant 1 - (t_1 - t_0) \cdot e^{-\varepsilon n}.$$

Proof. Let $\ell := \frac{n}{10} + n \cdot (2C+1)/w_+$. For any round $t \in [t_0, t_1]$, define

$$\mathcal{Q}^t := \Big\{ \Delta^t > C \cdot n \Big\} \cup \Big\{ \big| \{ s \in [t, t+\ell] \colon \delta^s \in (\varepsilon, 1-\varepsilon) \} \big| \geqslant \varepsilon \cdot n \Big\},$$

where $\varepsilon := \varepsilon(C)$ is the constant from Lemma 6.1. Note that event $\mathcal{Q}^t$ is logically equivalent to the statement: $\Delta^t \leqslant C \cdot n$ implies $|\{ s \in [t, t+\ell] \colon \delta^s \in (\varepsilon, 1-\varepsilon)\}| \geqslant \varepsilon \cdot n$. Then by Lemma 6.1,

$$\mathbf{Pr}\left[\mathcal{Q}^t \;\middle|\; \mathfrak{F}^{t_0} \right] \geqslant 1 - e^{-\varepsilon n},$$

and so the union bound gives

$$\mathbf{Pr}\left[\bigcap_{t=t_0}^{t_1} \mathcal{Q}^t \;\middle|\; \mathfrak{F}^{t_0} \right] \geqslant 1 - (t_1 - t_0) \cdot e^{-\varepsilon n}.$$

In the following, we will condition on the event $\bigcap_{t=t_0}^{t_1} \mathcal{Q}^t$. Next define for any $t \in [t_0, t_1 - \ell]$,

$$g(t) := \{ s \in [t, t+\ell] \colon \delta^s \in (\varepsilon, 1-\varepsilon) \}.$$

Then,

$$G_{t_0}^{t_1} \geqslant \left| \bigcup_{t=t_0}^{t_1 - \ell} g(t) \right| \geqslant \frac{\sum_{t=t_0}^{t_1-\ell} |g(t)|}{\max_{r \in [t, t+\ell]} \Big| \{ t \in [t_0, t_1 - \ell] \colon r \in g(t) \} \Big|} \geqslant \frac{\widetilde{G}_{t_0}^{t_1 - \ell} \cdot \varepsilon \cdot n}{\ell} \geqslant \frac{\widetilde{G}_{t_0}^{t_1 - 5Cn} \cdot \varepsilon}{5C},$$

where the last step used $\ell \leqslant 5Cn$. This completes the proof. $\square$

We will now prove using Lemma 6.2 that when $\Lambda^{t_0} \leqslant \exp(2c_6 n \log n)$, half of the rounds $t \in [t_0, t_0 + n^3 \log^3 n]$ satisfy $\Delta^t \leqslant Cn$ with constant probability. We will use this lemma in the recovery phase (Lemma 9.5).

Lemma 9.3. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. Then for $\varepsilon := \varepsilon(C)$ from Lemma 6.1, $r = \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$, and for any rounds t_0 and t_1 with $t_1 \geqslant t_0 + n^3 \log^3 n$, we have*

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \geqslant \frac{1}{2} - (t_1 - t_0) \cdot e^{-\varepsilon n},$$

where $c_6 > 0$ is the constant from Lemma 8.10.

Proof. Using Lemma 6.2, we have for any $t \geqslant 0$,

$$\mathbf{E}[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,] \leqslant \Upsilon^t - \frac{c_1}{n} \cdot \Delta^t + c_2.$$

By taking the expectations on both sides, we get

$$\mathbf{E}[\, \Upsilon^{t+1} \mid \mathfrak{F}^{t_0} \,] = \mathbf{E}[\, \mathbf{E}[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,] \mid \mathfrak{F}^{t_0} \,] \leqslant \mathbf{E}[\, \Upsilon^t \mid \mathfrak{F}^{t_0} \,] - \frac{c_1}{n} \cdot \mathbf{E}[\, \Delta^t \mid \mathfrak{F}^{t_0} \,] + c_2.$$

Applying this to rounds $t_0, t_0+1, \ldots, t_1+1$, we get

$$\begin{aligned}
\mathbf{E}[\, \Upsilon^{t_0+1} \mid \mathfrak{F}^{t_0} \,] &\leqslant \mathbf{E}[\, \Upsilon^{t_0} \mid \mathfrak{F}^{t_0} \,] - \frac{c_1}{n} \cdot \mathbf{E}[\, \Delta^{t_0} \mid \mathfrak{F}^{t_0} \,] + c_2, \\
\mathbf{E}[\, \Upsilon^{t_0+2} \mid \mathfrak{F}^{t_0} \,] &\leqslant \mathbf{E}[\, \Upsilon^{t_0+1} \mid \mathfrak{F}^{t_0} \,] - \frac{c_1}{n} \cdot \mathbf{E}[\, \Delta^{t_0+1} \mid \mathfrak{F}^{t_0} \,] + c_2, \\
&\vdots \\
\mathbf{E}[\, \Upsilon^{t_1+1} \mid \mathfrak{F}^{t_0} \,] &\leqslant \mathbf{E}[\, \Upsilon^{t_1} \mid \mathfrak{F}^{t_0} \,] - \frac{c_1}{n} \cdot \mathbf{E}[\, \Delta^{t_1} \mid \mathfrak{F}^{t_0} \,] + c_2.
\end{aligned}$$

Hence by induction, and using $\mathbf{E}[\, \Upsilon^{t_0} \mid \mathfrak{F}^{t_0} \,] = \Upsilon^{t_0}$, we get

$$\mathbf{E}[\, \Upsilon^{t_1+1} \mid \mathfrak{F}^{t_0} \,] \leqslant \Upsilon^{t_0} - \frac{c_1}{n} \cdot \sum_{r=t_0}^{t_1} \mathbf{E}[\, \Delta^r \mid \mathfrak{F}^{t_0} \,] + c_2 \cdot (t_1 - t_0 + 1).$$

Since the left hand side is at least 0, rearranging the above inequality yields

$$\sum_{r=t_0}^{t_1} \mathbf{E}[\, \Delta^r \mid \mathfrak{F}^{t_0} \,] \leqslant \Upsilon^{t_0} \cdot \frac{n}{c_1} + c_2 \cdot (t_1 - t_0 + 1) \cdot \frac{n}{c_1}.$$

We will now make use of the condition $\Lambda^{t_0} \leqslant \exp(2c_6 n \log n)$, and conclude by the first statement of Lemma 8.1 that

$$\Upsilon^{t_0} \leqslant \alpha^{-2} \cdot n \cdot (\log \Lambda^{t_0})^2 \leqslant 4 \cdot \alpha^{-2} \cdot c_6^2 \cdot n^3 \cdot \log^2 n.$$

So, we obtain that

$$\begin{aligned}
\sum_{r=t_0}^{t_1} \mathbf{E}[\, \Delta^r \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \,] &\leqslant 4 \cdot \alpha^{-2} \cdot c_6^2 \cdot n^3 \cdot \log^2 n \cdot \frac{n}{c_1} + c_2 \cdot (t_1 - t_0 + 1) \cdot \frac{n}{c_1} \\
&\leqslant 2 \cdot \frac{c_2}{c_1} \cdot (t_1 - t_0) \cdot n \leqslant C \cdot (t_1 - t_0) \cdot \frac{n}{c_1},
\end{aligned}$$

where we have used that $t_1 - t_0 \geqslant n^3 \log^3 n > 4 \cdot \alpha^{-2} \cdot c_6^2 \cdot n^3 \cdot \log^2 n + \frac{c_2}{c_1} \cdot n$ for constant $\alpha > 0$.

Finally, by applying Markov's inequality, we obtain

$$\mathbf{Pr}\left[\sum_{r=t_0}^{t_1} \Delta^r \leqslant C \cdot (t_1 - t_0) \cdot n/2 \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \leqslant \frac{1}{2},$$

and so at least half of the rounds $t \in [t_0, t_1]$ satisfy $\Delta^t \leqslant Cn$ w.p. at least $1/2$, i.e.,

$$\mathbf{Pr}\left[\widetilde{G}_{t_0}^{t_1} > \frac{1}{2} \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \geqslant \frac{1}{2}, \tag{9.3}$$

By Lemma 9.2, there exists a constant $\varepsilon := \varepsilon(C) > 0$

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > \frac{\varepsilon}{5C} \widetilde{G}_{t_0}^{t_1 - 5Cn} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \geqslant 1 - (t_1 - t_0) \cdot e^{-\varepsilon n}. \tag{9.4}$$

Then, noting that $\widetilde{G}_{t_0}^{t_1-5Cn} \geqslant \widetilde{G}_{t_0}^{t_1} - 5Cn$ and by taking the union bound of (9.3) and (9.4) we have,

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > \frac{\varepsilon}{5C} \cdot \left(\frac{1}{2} \cdot (t_1 - t_0) - 5Cn \right) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \geqslant \frac{1}{2} - (t_1 - t_0) \cdot e^{-\varepsilon n}.$$

Finally, since $\frac{t_1 - t_0}{4} > 5Cn$, we can deduce for $r = \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$,

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \exp(2c_6 n \log n) \right] \geqslant \frac{1}{2} - (t_1 - t_0) \cdot e^{-\varepsilon n}. \qquad \square$$

The next lemma establishes the key fact that w.h.p. when $\Lambda^{t_0} < \kappa_1 n$ there are many rounds close to t_0 with small absolute value potential. In contrast to Lemma 9.3, this claim is proven w.h.p. and will be used in the stabilization phase (Lemma 9.7).

Lemma 9.4. *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. Let $\varepsilon := \varepsilon(C)$ be as in Lemma 6.1, $r := \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$ and $\alpha > 0$ be any constant. Then, for any constants $\kappa_1, \kappa_2 > 0$ and for any rounds t_0 and t_1 satisfying $t_1 := t_0 + \kappa_2 \cdot n \log n$, we have*

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \right] \geqslant 1 - 3 \cdot n^{-12}.$$

Proof. For the constants c_1, c_2 given in Lemma 6.2, we define for any $t \geqslant t_0$ the sequence

$$Z^t := \Upsilon^t - c_2 \cdot (t - t_0) + \frac{c_1}{n} \sum_{i=t_0}^{t-1} \Delta^i.$$

This sequence forms a super-martingale since by Lemma 6.2,

$$\begin{aligned}
\mathbf{E}\left[Z^{t+1} \mid \mathfrak{F}^t \right] &= \mathbf{E}\left[\Upsilon^{t+1} - c_2 \cdot (t - t_0 + 1) + \frac{c_1}{n} \sum_{i=t_0}^{t} \Delta^i \;\middle|\; \mathfrak{F}^t \right] \\
&= \mathbf{E}\left[\Upsilon^{t+1} \mid \mathfrak{F}^t \right] - c_2 \cdot (t - t_0 + 1) + \frac{c_1}{n} \sum_{i=t_0}^{t} \Delta^i \\
&\leqslant \Upsilon^t + c_2 - \frac{c_1}{n} \cdot \Delta^t - c_2 \cdot (t - t_0 + 1) + \frac{c_1}{n} \sum_{i=t_0}^{t} \Delta^i \\
&= \Upsilon^t - c_2 \cdot (t - t_0) + \frac{c_1}{n} \sum_{i=t_0}^{t-1} \Delta^i \\
&= Z^t.
\end{aligned}$$

Further, let $\tau := \min\{t \geqslant t_0 \colon \max_{i \in [n]} |y_i^t| > \log^2 n\}$ and consider the stopped random variable

$$\widetilde{Z}^t := Z^{t \wedge \tau},$$

which is then also a super-martingale. Lemma 8.7 implies that

$$\mathbf{Pr}\left[\tau \leqslant t_1 \right] \leqslant n^{-12}, \tag{9.5}$$

i.e., the gap does not increase above $\log^2 n$ in any of the rounds $[t_0, t_1]$ w.h.p.

To prove concentration of $\widetilde{Z}^{t+1}$, we will now derive an upper bound on $\left| \widetilde{Z}^{t+1} - \widetilde{Z}^t \right|$ conditional on $\mathfrak{F}^t$ and $\Lambda^{t_0} \leqslant \kappa_1 \cdot n$.

Case 1 [$t \geqslant \tau$]. In this case, $\widetilde{Z}^{t+1} = Z^{(t+1) \wedge \tau} = Z^\tau$, and similarly, $\widetilde{Z}^t = Z^{t \wedge \tau} = Z^\tau$, so $|\widetilde{Z}^{t+1} - \widetilde{Z}^t| = 0$.

Case 2 $[t < \tau]$. Hence for t we have $\max_{i\in[n]} |y_i^t| < \log^2 n$ and thus Lemma 8.2 implies that the biggest change in the quadratic potential is bounded by $4w_- \cdot \log^2 n + 2w_-^2 \leqslant 10w_- \cdot \log^2 n$. Combining the two cases above, we conclude

$$|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant 10w_- \cdot \log^2 n.$$

Using Azuma's inequality (Lemma A.9) for super-martingales,

$$\begin{aligned}\mathbf{Pr}\left[\, \widetilde{Z}^{t_1+1} - \widetilde{Z}^{t_0} > \lambda \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \,\right] &\leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{t=t_0}^{t_1} (10w_- \cdot \log^2 n)^2}\right) \\ &= \exp\left(-\frac{\lambda^2}{(t_1 - t_0) \cdot 200w_-^2 \cdot \log^4 n}\right),\end{aligned}$$

which means that for $\lambda = n$ we conclude $\mathbf{Pr}\left[\, \widetilde{Z}^{t_1+1} > \widetilde{Z}^{t_0} + n \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \geqslant \kappa_1 \cdot n \,\right] \leqslant n^{-\omega(1)}$. Thus by (9.5) and the union bound we have

$$\mathbf{Pr}\left[\, Z^{t_1+1} \leqslant Z^{t_0} + n \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \geqslant \kappa_1 \cdot n \,\right] \geqslant 1 - 2 \cdot n^{-12}.$$

For the sake of a contradiction, assume now that at least half of the rounds $t \in [t_0, t_1]$ satisfy $\Delta^t \geqslant Cn$, which implies

$$\sum_{t=t_0}^{t_1} \Delta^t \geqslant \frac{t_1 - t_0 + 1}{2} \cdot C \cdot n.$$

If $Z^{t_1+1} \leqslant Z^{t_0} + n$ holds, then we have

$$\Upsilon^{t_1+1} - c_2 \cdot (t_1 - t_0 + 1) + \frac{c_1}{n} \sum_{t=t_0}^{t_1} \Delta^t \leqslant \Upsilon^{t_0} + n.$$

Rearranging the inequality above gives

$$\begin{aligned}\Upsilon^{t_1+1} &\leqslant \Upsilon^{t_0} + n + c_2 \cdot (t_1 - t_0 + 1) - \frac{c_1}{n} \sum_{t=t_0}^{t_1} \Delta^t \\ &\leqslant \Upsilon^{t_0} + n + c_2 \cdot (t_1 - t_0 + 1) - \frac{c_1}{n} \cdot \frac{t_1 - t_0 + 1}{2} \cdot C \cdot n \\ &\leqslant \Upsilon^{t_0} + n + (t_1 - t_0 + 1) \cdot (c_2 - \frac{c_1}{2} \cdot C).\end{aligned} \tag{9.6}$$

Recall that we start from a round t_0 where $\Lambda^{t_0} \leqslant \kappa_1 \cdot n$, and therefore also $\Upsilon^{t_0} \leqslant \left(\frac{4}{\alpha} \cdot \log \frac{4}{\alpha}\right)^2 \cdot \kappa_1 \cdot n$ by Lemma 8.1. Thus, recalling that $C > \frac{2c_2}{c_1}$, by (9.6) we have

$$\Upsilon^{t_1+1} \leqslant \left(\frac{4}{\alpha} \cdot \log \frac{4}{\alpha}\right)^2 \cdot \kappa_1 \cdot n + n + (\kappa_2 n \log n + 1) \cdot (c_2 - \frac{c_1}{2} \cdot C) < 0$$

which is a contradiction for large n. We conclude that if $Z^{t_1+1} \leqslant Z^{t_0} + n$, then half of the rounds $t \in [t_0, t_1]$ satisfy $\Delta^t \leqslant Cn$, thus this event holds w.p. $1 - 2 \cdot n^{-12}$, i.e.

$$\mathbf{Pr}\left[\, \widetilde{G}_{t_0}^{t_1} \geqslant \frac{1}{2} \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \,\right] \geqslant 1 - 2 \cdot n^{-12}. \tag{9.7}$$

Applying Lemma 9.2 gives

$$\mathbf{Pr}\left[\, G_{t_0}^{t_1} > \frac{\varepsilon}{5C} \cdot \widetilde{G}_{t_0}^{t_1 - 5Cn} \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \,\right] \geqslant 1 - (t_1 - t_0) \cdot e^{-\varepsilon n}. \tag{9.8}$$

Then, noting that $\widetilde{G}_{t_0}^{t_1-5Cn} \geqslant \widetilde{G}_{t_0}^{t_1} - 5Cn$ and by taking the union bound of (9.7) and (9.8) we have,

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > \frac{\varepsilon}{5C} \cdot \Big(\frac{1}{2} \cdot (t_1 - t_0) - 5Cn \Big) \;\Big|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \right] \geqslant 1 - (t_1 - t_0) \cdot e^{-\varepsilon n} - 2n^{-12}.$$

Since, $\frac{1}{4} \cdot (t_1 - t_0) = \frac{c_s}{4} \cdot n \log n > 5Cn$, we can deduce that for $r = \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$,

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\Big|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \kappa_1 \cdot n \right] \geqslant 1 - 3 \cdot n^{-12}. \qquad \square$$

9.3 Recovery of the Process

In the following, we start the analysis at round $t_0 = m - \Theta(n^3 \log^4 n)$ and using Lemma 8.10, we obtain that w.h.p. $\mathrm{Gap}(t_0) = \mathcal{O}(n \log n)$ and $\Lambda^t \leqslant \exp(2c_6 n \log n)$. Using Lemma 9.3 we conclude that with constant probability for a constant fraction of the rounds $t \in [t_0, t_0 + \Theta(n^3 \log^3 n)]$, it holds that $\delta^t \in (\varepsilon, 1-\varepsilon)$. Then we exploit the drop of the exponential potential function in those rounds to infer that with constant probability there is a round $s \in [t_0, t_0 + \Theta(n^3 \log^3 n)]$ with $\Lambda^s = \mathcal{O}(n)$. Then, we use a retry argument $\Theta(\log n)$ times, to amplify the probability of finding such a round. For each failure, we can still restart from a round with $\Lambda^t \leqslant \exp(2c_6 n \log n)$. Hence after these $\Theta(\log n)$ repetitions, we deduce that w.h.p., there exists a round $s \in [t_0, m]$, where $\Lambda^s \leqslant cn$, which yields Lemma 9.5. From that point onwards, we will use the stabilization lemma (Lemma 9.8) for the remaining $\mathcal{O}(n^3 \log^4 n)$ rounds (see Section 9.4).

Lemma 9.5 (Recovery). *Consider any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process. Then, for some constant $\alpha > 0$, and for $c > 1$ being the constant in Corollary 8.6, for any $m \geqslant 40n^3 \log^4 n$,*

$$\mathbf{Pr}\left[\bigcup_{s \in [m - 40n^3 \log^4 n, m]} \{\Lambda^s < cn\} \right] \geqslant 1 - n^{-10}.$$

Proof. Let $k := 40 \log n$ and $T := n^3 \log^3 n$. Consider $t_0 := m - k \cdot T$, and define k rounds $t_1 := t_0 + T, t_2 := t_0 + 2T, \ldots, t_k := m$. We now recall for any $0 \leqslant i \leqslant k$ the event

$$\mathcal{E}_{t_0}^{t_i} := \bigcap_{s \in [t_0, t_i]} \{\Lambda^s \geqslant cn\}.$$

Further, we define for any integer $0 \leqslant i \leqslant k$ the event

$$\mathcal{H}^{t_i} := \left\{ \Lambda^{t_i} \leqslant \exp(2c_6 n \log n) \right\}.$$

We will prove that, for any integer $0 \leqslant i \leqslant k$, we have

$$\mathbf{Pr}\left[\mathcal{E}_{t_0}^{t_i} \right] \leqslant (3/4)^i + 4 \cdot n^{-12}.$$

Thus for $i = k$, we obtain the statement of the lemma. We next establish the following claim:

Claim 9.6. *For any integer $1 \leqslant i \leqslant k$, we have $\mathbf{Pr}\left[\mathcal{E}_{t_0}^{t_i} \mid \mathfrak{F}^{t_{i-1}}, \mathcal{H}^{t_{i-1}} \right] \leqslant 3/4$.*

Proof of Claim. Using Lemma 9.3 for t_{i-1} and t_i (since $t_i - t_{i-1} = T$), we get that for $r = \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$,

$$\mathbf{Pr}\left[G_{t_{i-1}}^{t_i} > r \cdot T \;\Big|\; \mathfrak{F}^{t_{i-1}}, \mathcal{H}^{t_{i-1}} \right] \geqslant 1/2 - T \cdot e^{-\varepsilon n}. \tag{9.9}$$

Let $\gamma := \gamma(\varepsilon) = \frac{r}{2(1-r)} < 1$ as $r \leqslant 1/2$, and choose the constant $\alpha := \alpha(\varepsilon)$ so as to satisfy $\alpha \leqslant c_3 \gamma / (3w_- \cdot e^{2w_-})$, where $c_3 := c_3(\varepsilon) > 0$ is given by Lemma 8.4, and to satisfy either (8.1) or (8.2) depending on whether the process satisfies $\mathcal{W}_3$ or $\mathcal{P}_3$ respectively. Since all the previous steps (and constants determined by them) hold for any c (and the constants $C, \varepsilon, r, \gamma, \alpha$ and Lemma 8.4 do not depend on c) we can now set

$$c_s := \frac{2 \cdot 9}{c_3 \cdot \alpha \cdot r} > 0 \qquad \text{and} \qquad c := \max\left(\frac{3(c_3 + w_-) \cdot e^{2w_-}}{\alpha c_3}, \frac{2}{\alpha^2}\right) > 1,$$

so c satisfies the assumptions of Corollary 8.6. As α satisfies Lemma 9.1, $(\widetilde{\Lambda}_{t_{i-1}}^{t})_{t \geqslant t_{i-1}}$ is a super-martingale. Thus, $\mathbf{E}[\,\widetilde{\Lambda}_{t_{i-1}}^{t_i+1} \mid \mathfrak{F}_{t_{i-1}}\,] \leqslant \widetilde{\Lambda}_{t_{i-1}}^{t_{i-1}}$. Applying Markov's inequality gives

$$\mathbf{Pr}\left[\widetilde{\Lambda}_{t_{i-1}}^{t_i+1} > 5 \cdot \widetilde{\Lambda}_{t_{i-1}}^{t_{i-1}} \mid \mathfrak{F}^{t_{i-1}}, \mathcal{H}^{t_{i-1}}\right] \leqslant 1/5. \tag{9.10}$$

Thus by the definition of $\widetilde{\Lambda}_{t_{i-1}}^{t}$ and taking the union bound of (9.10) and (9.9), we have

$$\begin{aligned}\mathbf{Pr}\Bigg[\left\{\Lambda^{t_i+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_i}} \leqslant 5 \cdot \Lambda^{t_{i-1}} \cdot e^{\frac{c_3\alpha\gamma}{n} \cdot B_{t_{i-1}}^{t_i} - \frac{c_3\alpha}{n} \cdot G_{t_{i-1}}^{t_i}}\right\} \bigcap \left\{G_{t_{i-1}}^{t_i} \geqslant r \cdot T\right\} \;\Bigg|\; \mathfrak{F}^{t_{i-1}}, \mathcal{H}^{t_{i-1}}\Bigg]& \\ \geqslant 1 - 1/2 - o(1) - 1/5 \geqslant 1/4.&\end{aligned} \tag{9.11}$$

However, if the event $G_{t_{i-1}}^{t_i} \geqslant r \cdot T$ occurs, then

$$\Lambda^{t_i+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_i}} \leqslant 5 \cdot \Lambda^{t_{i-1}} \cdot \exp\left(\frac{c_3\alpha\gamma}{n} \cdot B_{t_{i-1}}^{t_i} - \frac{c_3\alpha}{n} \cdot G_{t_{i-1}}^{t_i}\right) \leqslant 5 \cdot \Lambda^{t_i} \cdot \exp\left(-\frac{c_3\alpha}{n} \cdot \frac{r}{2} \cdot T\right),$$

where the second inequality used $\gamma = \frac{r}{2(1-r)}$. Further, since we condition on $\mathcal{H}^{t_i}$, we obtain

$$\Lambda^{t_i+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_i}} \leqslant 5 \cdot \exp\left(2c_6 n \log n\right) \cdot \exp\left(-\frac{c_3\alpha}{n} \cdot \frac{r}{2} \cdot \left(n^3 \log^3 n\right)\right) < 1,$$

as $T = n^3 \log^3 n$ for any $i \leqslant k$. Recall that $\Lambda^t \geqslant n$ holds deterministically for any $t \geqslant 0$, thus we have a contradiction. We conclude that the event $\neg\mathcal{E}_{t_0}^{t_i} = \{\bigcup_{r \in [t_0, t_i]} \Lambda^r < cn\}$ is implied whenever the events $\{G_{t_{i-1}}^{t_i} \geqslant r \cdot T\}$, $\{\widetilde{\Lambda}_{t_{i-1}}^{t_i+1} \leqslant 5 \cdot \widetilde{\Lambda}_{t_{i-1}}^{t_{i-1}}\}$, and $\mathcal{H}^{t_{i-1}}$ all occur. The bound on $\mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i} \mid \mathfrak{F}^{t_{i-1}}, \mathcal{H}^{t_{i-1}}\,\right] = 1 - \mathbf{Pr}\left[\,\neg\mathcal{E}_{t_0}^{t_i} \mid \mathcal{H}^{t_{i-1}}, \mathfrak{F}^{t_{i-1}}\,\right]$ then follows from (9.11). $\Diamond$

By the second statement in Lemma 8.10, we get that for any $0 \leqslant i \leqslant k$

$$\mathbf{Pr}\left[\,\neg\mathcal{H}^{t_i}\,\right] \leqslant n^{-12}.$$

Combining this bound with Claim 9.6, for any $1 \leqslant i \leqslant k$ we have

$$\begin{aligned}\mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i}\,\right] &= \mathbf{Pr}\left[\,\mathcal{H}^{t_{i-1}} \cap \mathcal{E}_{t_0}^{t_i}\,\right] + \mathbf{Pr}\left[\,\neg\mathcal{H}^{t_{i-1}} \cap \mathcal{E}_{t_0}^{t_i}\,\right] \\ &\leqslant \mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i} \mid \mathcal{H}^{t_{i-1}}, \mathcal{E}_{t_0}^{t_{i-1}}\,\right] \cdot \mathbf{Pr}\left[\,\mathcal{H}^{t_{i-1}} \cap \mathcal{E}_{t_0}^{t_{i-1}}\,\right] + \mathbf{Pr}\left[\,\neg\mathcal{H}^{t_{i-1}}\,\right] \\ &\leqslant \mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i} \mid \mathcal{H}^{t_{i-1}}, \mathcal{E}_{t_0}^{t_{i-1}}\,\right] \cdot \mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_{i-1}}\,\right] + n^{-12} \\ &\leqslant (3/4) \cdot \mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_{i-1}}\,\right] + n^{-12}.\end{aligned}$$

By the second statement of Lemma A.12 for $(\mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i}\,\right])_{i \geqslant 0}$, $a := 3/4 < 1$ and $b := n^{-12}$, we get

$$\mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_i}\,\right] \leqslant \mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_0}\,\right] \cdot (3/4)^i + \frac{n^{-12}}{1 - \frac{3}{4}} \leqslant (3/4)^i + 4 \cdot n^{-12},$$

as $\mathbf{Pr}\left[\,\mathcal{E}_{t_0}^{t_0}\,\right] \leqslant 1$. Setting $i := k = 40 \log n$, gives the result since $(3/4)^{40 \log n} + 4 \cdot n^{-12} \leqslant n^{-10}$. $\square$

9.4 Stabilization of the Process

The next lemma establishes that a small value of the exponential potential function is preserved for some longer time period. We will call this property of the exponential potential function of being "trapped" in some region *stabilization.* More precisely, we prove in the lemma below that if for some round t_0 the exponential potential is not too small, i.e., at most $2cn$, then within the next $\mathcal{O}(n \log n)$ rounds, the exponential potential will be smaller than cn at least once w.h.p..

Lemma 9.7 (Stabilization). *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process there exists constants $\alpha \in (0, 1/w_-]$, $c > 1$ and $c_s > 0$, such that for any round $t_0 \geqslant 0$,*

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0 + c_s n \log n - 1]} \{\Lambda^t < cn\} \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \in [cn, 2cn] \right] \geqslant 1 - \frac{1}{2} \cdot n^{-7}.$$

Proof. Let $t_1 := t_0 + c_s n \log n$, for c_s a constant to be defined below. By Lemma 9.4 with $\kappa_1 = 2c$ and $\kappa_2 = c_s$ we have for $r := \min\{\frac{\varepsilon}{20C}, \frac{1}{2}\}$ (the same r as defined in the proof of Lemma 9.5),

$$\mathbf{Pr}\left[G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \in [cn, 2cn] \right] \geqslant 1 - 3 \cdot n^{-12}. \tag{9.12}$$

Again, we use $\gamma = \frac{r}{2(1-r)} \leqslant 1$ and choose α and c as in Lemma 9.5. As α satisfies Lemma 9.1, $(\widetilde{\Lambda}_{t_{i-1}}^t)_{t \geqslant t_{i-1}}$ is a super-martingale, so $\mathbf{E}[\, \widetilde{\Lambda}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0} \,] \leqslant \widetilde{\Lambda}_{t_0}^{t_0} = \Lambda^{t_0}$. Hence, using Markov's inequality we get $\mathbf{Pr}\left[\widetilde{\Lambda}_{t_0}^{t_1} > \Lambda^{t_0} \cdot n^8 \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \in [cn, 2cn] \right] \leqslant n^{-8}$. Thus, by the definition of $\widetilde{\Lambda}_{t_0}^t$ given in Lemma 9.1, we have

$$\mathbf{Pr}\left[\Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1 - 1}} \leqslant \Lambda^{t_0} \cdot n^8 \cdot \exp\left(\frac{c_3 \alpha \gamma}{n} \cdot B_{t_0}^{t_1 - 1} - \frac{c_3 \alpha}{n} \cdot G_{t_0}^{t_1 - 1} \right) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \in [cn, 2cn] \right] \geqslant 1 - n^{-8}. \tag{9.13}$$

Further, if in addition to the two events $\{\widetilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot n^8\}$ and $\{\Lambda^{t_0} \in [cn, 2cn]\}$, also the event $\{G_{t_0}^{t_1 - 1} \geqslant r \cdot (t_1 - t_0)\}$ holds, then

$$\begin{aligned}
\Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1 - 1}} &\leqslant \Lambda^{t_0} \cdot n^8 \cdot \exp\left(\frac{c_3 \alpha \gamma}{n} \cdot B_{t_0}^{t_1 - 1} - \frac{c_3 \alpha}{n} \cdot G_{t_0}^{t_1 - 1} \right) \\
&\leqslant 2cn^9 \cdot \exp\left(\frac{c_3 \alpha \gamma}{n} \cdot (1 - r) \cdot (t_1 - t_0) - \frac{c_3 \alpha}{n} \cdot r \cdot (t_1 - t_0) \right) \\
&\overset{(a)}{=} 2cn^9 \cdot \exp\left(-\frac{c_3 \alpha}{n} \cdot \frac{r}{2} \cdot (t_1 - t_0) \right) \\
&= 2cn^9 \cdot \exp\left(-\frac{c_3 \alpha}{n} \cdot \frac{r}{2} \cdot \frac{2 \cdot 9}{c_3 \cdot \alpha \cdot r} \cdot n \log n \right) = 2c,
\end{aligned}$$

where (a) holds since $\gamma(1 - r) - r = -\frac{r}{2}$ due to the definition of γ. Observe that $\Lambda^{t_1} \geqslant n$ holds deterministically, so we deduce from the above inequality that $\mathbf{1}_{\mathcal{E}_{t_0}^{t_1 - 1}} = 0$, that is,

$$\mathbf{Pr}\left[\neg \mathcal{E}_{t_0}^{t_1 - 1} \;\middle|\; \mathfrak{F}^{t_0}, \quad \widetilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot n^8, \quad \Lambda^{t_0} \in [cn, 2cn], \quad G_{t_0}^{t_1 - 1} \geqslant r \cdot (t_1 - t_0) \right] = 1.$$

Recalling the definition of $\mathcal{E}_{t_0}^{t_1 - 1} = \bigcap_{r \in [t_0, t_1 - 1]} \{\Lambda^r \geqslant cn\}$, and taking the union bound over (9.12) and (9.13) yields

$$\mathbf{Pr}\left[\bigcup_{r \in [t_0, t_0 + c_s n \log n - 1]} \{\Lambda^r < cn\} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \in [cn, 2cn] \right] \geqslant 1 - 3 \cdot n^{-12} - n^{-8} \geqslant 1 - \frac{1}{2} \cdot n^{-7},$$

as claimed. $\square$

The next lemma shows that if there is a round with linear (exponential) potential then the gap is at most logarithmic for the next n^4 rounds. The idea is to repeatedly apply Lemma 9.7 to find a interval of length n^4 where any contiguous sub-interval of length $\Theta(n \log n)$ contains a round with linear potential. The result then follows since a linear potential implies a logarithmic gap, and the gap can grow by at most $\Theta(\log n)$ in $\Theta(n \log n)$ rounds.

Lemma 9.8. *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_2 \cap \mathcal{W}_3$-process there exists a constant $\kappa > 0$ such that, for any rounds t_0, t_1 with $t_0 < t_1 \leqslant t_0 + n^4$,*

$$\mathbf{Pr}\left[\max_{i\in[n]} |y_i^{t_1}| \leqslant \kappa \cdot \log n \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} < cn \right] \geqslant 1 - \frac{1}{2} \cdot n^{-3}.$$

Proof. Choose the constants $\alpha \in (0, 1/w_-]$, $c > 1$ and $c_s > 0$ so as to satisfy Lemma 9.7. Next define the event

$$\mathcal{M}_{t_0}^{t_1} = \left\{\text{for all } t \in [t_0, t_1] \text{ there exists } s \in [t, t + c_s n \log n] \text{ such that } \Lambda^s < cn\right\},$$

that is, if $\mathcal{M}_{t_0}^{t_1}$ holds then we have $\Lambda^s < cn$ at least once every $c_s n \log n$ rounds.

Assume now that $\mathcal{M}_{t_0}^{t_1}$ holds. Choosing $t = t_1$, there exists $s \in [t_1, t_1 + c_s n \log n]$ such that $\Lambda^s < cn$, which in turn implies by definition of Λ that $\max_{i\in[n]} |y_i^s| \leqslant \frac{1}{\alpha} \cdot \log(cn) < \frac{2}{\alpha} \cdot \log n$. Clearly, any y_i^t can decrease by at most w_-/n in each round, and from this it follows that if $\mathcal{M}_{t_0}^{t_1}$ holds, then $\max_{i\in[n]} y_i^{t_1} \leqslant \max_{i\in[n]} |y_i^s| + w_- \cdot c_s \log n \leqslant \kappa \cdot \log n$, for $\kappa := \frac{2}{\alpha} + w_- \cdot c_s$.

If $t_1 \geqslant t_0 + c_s n \log n$ and $\mathcal{M}_{t_0}^{t_1}$ holds, then choosing $s = t_1 - c_s n \log n$, there exists $s \in [t_1 - c_s n \log n, t_1]$ such that $\Lambda^s < cn$. (in case $t_1 - c_s n \log n < t_0$, then we arrive at the same conclusion by choosing $s = t_0$). This in turn implies $\max_{i\in[n]} |y_i^s| < \frac{2}{\alpha} \cdot \log n$. Hence

$$\min_{i\in[n]} y_i^{t_1} \geqslant \min_{i\in[n]} y_i^s - w_- \cdot c_s \log n \geqslant - \max_{i\in[n]} |y_i^s| - w_- \cdot c_s \log n \geqslant -\kappa \cdot \log n.$$

Hence, $\mathcal{M}_{t_0}^{t_1}$ (conditioned on $\Lambda^{t_0} < cn$) implies that $\max_{i\in[n]} |y_i^{t_1}| \leqslant \kappa \cdot \log n$. It remains to bound $\mathbf{Pr}\left[\neg \mathcal{M}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} < cn\right]$.

Note that since we start with $\Lambda^{t_0} < cn$, if for some round j, $\Lambda^j > 2cn$, there must exist some $s \in [t_0, j)$, such that $\Lambda^s \in [cn, 2cn]$, since for every $t \geqslant 0$ it holds $\Lambda^{t+1} \leqslant \Lambda^t \cdot e^{\alpha w_-} \leqslant 2\Lambda^t$ and $\alpha \leqslant 1/w_-$. Let $t_0 < r_1 < r_2 < \cdots$ and $t_0 =: s_0 < s_1 < \cdots$ be two interlaced sequences defined recursively for $i \geqslant 1$ by

$$r_i := \inf\{r > s_{i-1} : \Lambda^r \in [cn, 2cn]\} \qquad \text{and} \qquad s_i := \inf\{s > r_i : \Lambda^t < cn\}.$$

Thus we have

$$t_0 = s_0 < r_1 < s_1 < r_2 < s_2 < \cdots,$$

and since $r_i > r_{i-1}$ we have $r_{t_1 - t_0} \geqslant t_1 - t_0$. Observe that if the event $\cap_{i=1}^{t_1 - t_0} \{s_i - r_i \leqslant c_s n \log n\}$ holds, then also $\mathcal{M}_{t_0}^{t_1}$ holds.

Recall that by Lemma 9.7 we have for any $i = 1, 2, \ldots, t_1 - t_0$ and any $r = t_0 + 1, \ldots, t_1$

$$\mathbf{Pr}\left[\bigcup_{t \in [r_i, r_i + c_s n \log n - 1]} \{\Lambda^t < cn\} \;\middle|\; \mathfrak{F}^r,\ \Lambda^r \in [cn, 2cn], r_i = r \right] \geqslant 1 - \frac{1}{2} \cdot n^{-7},$$

and by negating and the definition of s_i,

$$\mathbf{Pr}\left[s_i - r_i > c_s n \log n \;\middle|\; \mathfrak{F}^r, \Lambda^r \in [cn, 2cn], r_i = r \right] \leqslant \frac{1}{2} n^{-7}.$$

Since the above bound holds for any i and $\mathfrak{F}^r$, with $r_i = r$, it follows by the union bound over all $i = 1, 2, \ldots, t_1 - t_0$, as $t_1 - t_0 \leqslant n^4$,

$$\mathbf{Pr}\left[\neg \mathcal{M}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} < cn\right] \leqslant (t_1 - t_0) \cdot \frac{1}{2} n^{-7} \leqslant \frac{1}{2} n^{-3}. \qquad \square$$

9.5 Completing the Proof of Theorem 4.15

We are now ready to complete the proof.

Theorem 4.15 (restated). *For any $\mathcal{P}_3 \cap \mathcal{W}_2$-process or $\mathcal{P}_3 \cap \mathcal{W}_3$-process, there exists a constant $\kappa > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\max_{i\in[n]}\left|x_i^m - \frac{W^t}{n}\right| \leqslant \kappa \log n\right] \geqslant 1 - n^{-3};$$

so in particular, $\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \log n\,\right] \geqslant 1 - n^{-3}$.

Proof. Consider an arbitrary round $m \geqslant 1$. If $m < 40n^3 \log^4 n$, then the claim follows by Lemma 9.8 since $\Lambda^{t_0} = n < cn$.

Otherwise, let $t_0 := m - 40n^3 \log^4 n$. Firstly, by Lemma 9.5, we get

$$\mathbf{Pr}\left[\bigcup_{s\in[t_0,m]} \{\Lambda^s \leqslant cn\}\right] \geqslant 1 - n^{-10}.$$

Hence for $\tau := \inf\{s \geqslant t_0 \colon \Lambda^s < cn\}$ we have $\mathbf{Pr}\left[\,\tau \leqslant m\,\right] \geqslant 1 - n^{-10}$.

Secondly, using Lemma 9.8, there is some constant $\kappa := \kappa(\alpha)$ such that for any round $s \in [t_0, m]$

$$\mathbf{Pr}\left[\max_{i\in[n]}|y_i^m| \leqslant \kappa \cdot \log n \;\middle|\; \mathfrak{F}^s, \Lambda^s < cn\right] \geqslant 1 - \frac{1}{2} \cdot n^{-3}.$$

Combining the two inequalities from above,

$$\begin{aligned}
\mathbf{Pr}\left[\max_{i\in[n]}|y_i^m| \leqslant \kappa \cdot \log n\right] &\geqslant \sum_{s=t_0}^{m} \mathbf{Pr}\left[\,\tau = s\,\right] \cdot \mathbf{Pr}\left[\max_{i\in[n]}|y_i^m| \leqslant \kappa \cdot \log n \;\middle|\; \tau = s\right] \\
&\geqslant \sum_{s=t_0}^{m} \mathbf{Pr}\left[\,\tau = s\,\right] \cdot \mathbf{Pr}\left[\max_{i\in[n]}|y_i^m| \leqslant \kappa \cdot \log n \;\middle|\; \mathfrak{F}^s, \Lambda^s < cn\right] \\
&\geqslant \left(1 - \frac{1}{2}n^{-3}\right) \cdot \mathbf{Pr}\left[\,\tau \leqslant m\,\right] \\
&\geqslant \left(1 - \frac{1}{2}n^{-3}\right) \cdot \left(1 - n^{-10}\right) \geqslant 1 - n^{-3}.
\end{aligned}$$

□

10 Lower Bounds

In this section we shall prove several lower bounds for both filling and non-filling processes. See Table 1 for an concise overview of our lower bounds and together with previously known lower bounds. One interesting aspect is that our lower bound for non-filling processes makes use of the quantile stabilization result from the previous section (Lemma 9.4).

10.1 Lower Bounds for Uniform Processes

We first prove a general lower bound which holds for any process that picks bins for allocation uniformly and can then increment the load of that chosen bin arbitrarily. Since Twinning and Packing choose a uniform bin for allocation at each round, the result below immediately yields a gap bound of $\Omega(\frac{\log n}{\log \log n})$ for these processes for $m = \mathcal{O}(n)$ rounds.

Lemma 10.1. *Consider any allocation process, which at each round $t \geqslant 0$, picks a bin i^t uniformly. Furthermore, assume that at any round $t \geqslant 0$ the allocation process increments the* *load of bin i^t by some function $f^t \geqslant 1$, which may depend on $\mathfrak{F}^t$ and i^t. Then, there is a constant $d_1 > 0$ such that*

$$\mathbf{Pr}\left[\mathrm{Gap}\left(\frac{n}{2}\right) \geqslant d_1 \cdot \frac{\log n}{\log\log n} \right] \geqslant 1 - o(1).$$

Proof. Recall the fact (see, e.g., [25]) that in a One-Choice process with $n/2$ balls into n bins, with probability at least $1 - o(1)$ there is a bin $i \in [n]$ which will be chosen at least $\kappa \frac{\log n}{\log\log n}$ times during the first $n/2$ allocations, where $\kappa > 0$ is some constant. Hence in our process

$$x_i^{n/2} \geqslant \left(\kappa \frac{\log n}{\log\log n} \right) \cdot 1,$$

for some $i \in [n]$ w.h.p. as at least one ball is allocated at each round.

Let us define $f^* := \max_{1 \leqslant t \leqslant n/2} f^t$ to be the largest number of balls allocated in one round. Then, clearly, $W^{n/2} \leqslant (n/2) \cdot f^*$. Furthermore, there must be at least one bin $j \in [n]$ which receives f^* balls in one of the first $n/2$ rounds. Thus for any such bin,

$$x_j^{n/2} \geqslant f^*,$$

and therefore the gap is lower bounded by

$$\mathrm{Gap}(n/2) \geqslant \max\left\{ x_i^{n/2}, x_j^{n/2} \right\} - \frac{W^{n/2}}{n} \geqslant \max\left\{ \kappa \frac{\log n}{\log\log n}, f^* \right\} - \frac{f^*}{2}.$$

If $f^* \geqslant \kappa \frac{\log n}{\log\log n}$, then the lower bound is $\frac{1}{2} f^* \geqslant \frac{\kappa}{2} \frac{\log n}{\log\log n}$. Otherwise, $f^* < \kappa \frac{\log n}{\log\log n}$, and the lower bound is at least $\kappa \frac{\log n}{\log\log n} - \frac{\kappa}{2} \frac{\log n}{\log\log n} = \frac{\kappa}{2} \frac{\log n}{\log\log n}$. □

The technique used in the proof above can be extended to the special case of Packing, yielding a more general lower bound which holds for any round $m \geqslant n$.

Lemma 10.2. *For Packing, there is a constant $d_2 > 0$ such that for any round $m \geqslant n$ we have*

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \geqslant d_2 \cdot \frac{\log n}{\log\log n} \right] \geqslant 5/8.$$

Proof. Consider the round $t_0 := m - n$ and the interval $[t_0, m]$. For the Packing process, it then follows from (5.13) that there is a constant $c_6 > 0$ such that

$$\mathbf{E}\left[\Phi^{t_0} \right] \leqslant c_6 \cdot n,$$

where the constant α in $\Phi^{t_0} = \sum_{i\colon y_i^t \geqslant 2} \exp\left(\alpha y_i^{t_0}\right)$ is given by $\alpha = \min\{1/101, (1/40) \cdot c_2/c_1\}$ for the constants c_1, c_2 are given by Lemma 5.2. Hence by Markov's inequality,

$$\mathbf{Pr}\left[\Phi^{t_0} \geqslant 8 c_6 n \right] \leqslant \frac{1}{8}. \tag{10.1}$$

Conditioning on $\Phi^{t_0} \geqslant 8c_6 n$ implies

$$\Delta^{t_0} = 2 \cdot \sum_{i\colon y_i^{t_0} \geqslant 0} y_i^{t_0} \leqslant 8 \cdot n + \frac{1}{\alpha} \sum_{i\colon y_i^{t_0} \geqslant 2} \exp\left(\alpha y_i^t\right) \leqslant \kappa_1 \cdot n,$$

for some constant $\kappa_1 > 0$, where we used the fact that $z \leqslant \exp(z)$ for any $z \geqslant 0$.

Next define $\mathcal{B} := \left\{ i \in [n] \colon y_i^t \geqslant -2\kappa_1 \right\}$. Then, as $\Delta^{t_0} \leqslant \kappa_1 \cdot n$, it follows that $|\mathcal{B}| \geqslant n/2$. Further, using Claim 5.9, we conclude that for any round $s \geqslant 0$, $\Delta^{s+1} \leqslant \Delta^s + 4$ and thus

$$\Delta^t \leqslant \Delta^{t_0} + 4(t - t_0),$$

Hence
$$\sum_{t=t_0}^{m} \Delta^t \leqslant \kappa_2 \cdot n^2,$$
holds deterministically for some constant κ_2, where we can assume $\kappa_2 > 1$. Next note that the expected total number of balls allocated in a round $t \in [t_0, m]$, denoted by w^t, satisfies
$$w^t = \sum_{i\colon y_i^t \geqslant 0} \frac{1}{n} \cdot 1 + \sum_{i\colon y_i^t < 0} \frac{1}{n} \cdot \left(\lceil -y_i^t \rceil + 1\right) \leqslant \frac{1}{n} \cdot \left(\Delta^t + n\right),$$
as $\sum_{i\colon y_i^t<0} -y_i^t = \frac{1}{2}\Delta^t$. Thus the expected number of balls allocated overall in all rounds $t \in [t_0, m]$ together satisfies
$$\mathbf{E}\left[\sum_{t=t_0}^{m} w^t \,\middle|\, \mathfrak{F}^{t_0}, \Phi^{t_0} < 8c_6 n \right] \leqslant 2\kappa_2 \cdot n.$$
Using Markov's inequality,
$$\mathbf{Pr}\left[\sum_{t=t_0}^{m} w^t \geqslant 16\kappa_2 \cdot n \,\middle|\, \mathfrak{F}^{t_0}, \Phi^{t_0} < 8c_6 n \right] \leqslant \frac{1}{8}. \qquad (10.2)$$
In the following, let us also condition on the event $\sum_{t=t_0}^{m} W^t < 16\kappa_2 \cdot n$, which implies that the average load increases by at most $16\kappa_2$ between rounds t_0 and $m = t_0 + n$.

Next consider the allocation to the bins in $\mathcal{B}$ during the time-interval $[t_0, m]$. The maximum number of times a bin in $\mathcal{B}$ is chosen corresponds to the maximum load of a One-Choice with (at least) $n/2$ balls into $n/2$ bins, which is $\Omega(\frac{\log n}{\log\log n})$ with probability at least $1 - o(1)$ (see, e.g., Raab and Steger [25]). Every time such a bin in $\mathcal{B}$ is chosen, its load is incremented by at least one, so
$$\mathbf{Pr}\left[\max_{i \in \mathcal{B}} (x_i^t - x_i^{t_0}) \geqslant \kappa_3 \cdot \frac{\log n}{\log\log n} \,\middle|\, \mathfrak{F}^{t_0}, \Phi^{t_0} < 8c_6 n \right] \geqslant 1 - o(1), \qquad (10.3)$$
for some suitable constant $\kappa_3 > 0$. Finally, taking the union bound over (10.1), (10.2) and (10.3), we conclude that with probability at least $1 - 1/8 - 1/8 - o(1) \geqslant 5/8$ it holds that
$$\begin{aligned}
\max_{i \in [n]} y_i^m &\geqslant \max_{i \in \mathcal{B}} \left(x_i^m - \frac{W^m}{n} \right) \\
&= \max_{i \in \mathcal{B}} \left(x_i^m - x_i^{t_0} + x_i^{t_0} - \frac{W^{t_0}}{n} + \frac{W^{t_0}}{n} - \frac{W^m}{n} \right) \\
&\geqslant \max_{i \in \mathcal{B}} \left(x_i^m - x_i^{t_0} \right) + \min_{i \in \mathcal{B}} \left(x_i^{t_0} - \frac{W^{t_0}}{n} \right) + \left(\frac{W^{t_0}}{n} - \frac{W^m}{n} \right) \\
&\geqslant \kappa_3 \cdot \frac{\log n}{\log\log n} - 2\kappa_1 - 16\kappa_2,
\end{aligned}$$
and the claim follows for some constant $d_2 > 0$ since κ_1, κ_2 and κ_3 are all positive constants. □

10.2 Lower Bounds for Non-Filling Processes

We shall now define a new condition for allocation processes which is satisfied for many natural processes, including Twinning, Thinning, and $(1 + \beta)$ with constant β.

- **Condition** $\mathcal{P}_4$: for any $\varepsilon > 0$ there exists a constant $0 < k_3 \leqslant 1$ such that for all $t \geqslant 0$ with $\delta^t \in (\varepsilon, 1 - \varepsilon)$ and all bins $i \in [n]$ we have
$$p_i^t \geqslant \frac{k_3}{n}.$$

So essentially this condition implies that in any round t where the mean quantile is in $(\varepsilon, 1-\varepsilon)$, there is at least a $\Omega(1/n)$-probability of allocating to each bin.

We shall now observe that

- For the Mean-Thinning process, the probability of allocating to an overloaded bin $i \in [n]$ is $p_i^t = \frac{\delta^t}{n}$. Thus condition $\mathcal{P}_4$ is satisfied with $k_3 := \varepsilon$.
- For the Twinning process, we have $p_i^t = \frac{1}{n}$. So $\mathcal{P}_4$ is satisfied with $k_3 := 1$.
- The $(1+\beta)$-process has $p_i^t = \frac{1-\beta}{n} + \frac{\beta(2i-1)}{n^2} > \frac{1-\beta}{n}$, satisfying $\mathcal{P}_4$ with $k_3 := 1-\beta$.

The next claim shows that for many rounds the mean quantile is not at the extremes.

Claim 10.3. *Consider any process satisfying the conditions $\mathcal{P}_2$ and $\mathcal{W}_3$, or, $\mathcal{W}_2$ and $\mathcal{P}_3$. For any $m = \Theta(n \log n)$ and $\varepsilon > 0$, let $G_1^m := G_1^m(\varepsilon)$ be the number of rounds $s \in [1, m]$ with $\delta^s \in (\varepsilon, 1-\varepsilon)$. Then there exists some $\varepsilon > 0$ and a constant $d_3 := d_3(\varepsilon) > 0$ such that*

$$\mathbf{Pr}\left[\, G_1^m \geqslant d_3 \cdot m \,\right] \geqslant 1 - 2 \cdot n^{-12}.$$

Proof. By applying Lemma 9.4 for $t_0 = 0$, (where $\Lambda^0 = n$), we get that there exists a constant $C > 0$ such that $\Delta^t \leqslant C \cdot n$ for a constant fraction of the rounds t in the range $[1, m]$ with probability at least $1 - n^{-12}$. The claim then follows by applying Lemma 6.1 at each of these rounds. $\square$

Lemma 10.4. *Consider any process satisfying $\mathcal{P}_4$ and either the conditions $\mathcal{P}_2$ and $\mathcal{W}_3$, or, $\mathcal{W}_2$ and $\mathcal{P}_3$. Let $k = k_3 d_3/1000$ where k_3 is specified by $\mathcal{P}_4$ and d_3 is from Claim 10.3. Then*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(k \cdot n \log n) \geqslant k \cdot \log n \,\right] \geqslant 1 - o(1).$$

Proof. Let $m = k \cdot n \log n$. By Claim 10.3 w.h.p. there exists some constants $\varepsilon > 0$ and $d_3 > 0$ such that there are at least $d_3 m$ rounds $s \in [0, m]$ with $\delta^t \in (\varepsilon, 1-\varepsilon)$. Denote this set of rounds by S and observe that by condition $\mathcal{P}_4$, for any $s \in S$ and $i \in [n]$ we have $p_i^s \geqslant k_3/n$.

Observe that we can couple the location of the balls allocated as rounds in S to locations under a one choice process as follows: before each round $s \in S$ we sample an independent Bernoulli random variable $X_s \sim \mathsf{Ber}(k_3)$ with success probability k_3. If $X_s = 1$ then we allocate the ball(s) to a uniformly random bin. Otherwise, if $X_s = 0$ we allocate the ball(s) to the i_{th} loaded bin with probability $(p_i^s - k_3/n)/(1-k_3)$. If we let $X = \sum_{s \in S} X_s$ then it follows that, conditional on $|S| \geqslant d_3 m$ we have $X > k_3 d_3 m/2$ w.h.p. by the Chernoff bound. It follows that w.h.p. at least $(k_3 d_3 k/2) \cdot n \log n$ balls are allocated according to the one choice protocol.

Raab and Steger [25] (see also [24, Section 4]) show that if $cn \log n$ balls are allocated according to the one choice protocol, for any constant $c > 0$, then w.h.p. the max load is at least $(c + \sqrt{c}/10) \log n$. Thus if we choose $k = k_3 d_3/1000$ then we have

$$\frac{\mathrm{Gap}(m)}{\log n} \geqslant \frac{k_3 d_3 k}{2} + \frac{1}{10}\sqrt{\frac{k_3 d_3 k}{2}} - k = \frac{k_3^2 d_3^2}{2000} + \frac{1}{10} \cdot \frac{k_3 d_3}{20\sqrt{5}} - \frac{k_3 d_3}{1000} > \frac{k_3 d_3}{1000},$$

with probability $1 - o(1)$ by taking the union bound of these three events. $\square$

Hence, we can deduce from the lemma above by recalling that Mean-Thinning, Twinning and $(1+\beta)$ all satisfy $\mathcal{P}_4$ and either the conditions $\mathcal{P}_2$ and $\mathcal{W}_3$, or, $\mathcal{W}_2$ and $\mathcal{P}_3$:

Corollary 10.5. *For either the* Mean-Thinning*,* Twinning *or $(1+\beta)$-processes, there exist a constant $k > 0$ (different for each process) such that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(k \cdot n \log n) \geqslant k \cdot \log n \,\right] \geqslant 1 - o(1).$$

11 Conclusions

In this work we introduced two general frameworks that imply a $\mathcal{O}(\log n)$ gap bound via conditions on the probability bias or weight/filling bias. This framework can be applied to a variety of known but also new allocation processes to deduce a tight or nearly-tight gap bound. One important novelty of our proof is to relate the exponential potential to the quadratic (and absolute value) potential in order to prove stabilization.

One natural open problem is to refine some of the bounds, e.g., formulate stronger conditions which imply a gap bound of $o(\log n)$. A different direction is to study other ranges of our conditions, e.g., in $\mathcal{W}_2$, we may even consider $w_+ = 0$ (overloaded bins are avoided at the cost of more bin samples), or $w_+ = -1$ (balls from overloaded bins are processed and removed).

A different direction is to investigate processes satisfying relaxed versions of the conditions, i.e., processes where the probability bias (or the weight bias) is close to 0 (as $n \to \infty$). In this case, one would expect the gap to grow by a function of the bias, and determining this exact relationship seems to be a challenging open problem.

Acknowledgments

We thank David Croydon and Martin Krejca for some helpful discussions.

References

[1] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. *SIAM J. Comput.*, 29(1):180–200, 1999.

[2] P. Berenbrink, A. Czumaj, A. Steger, and B. Vöcking. Balanced allocations: the heavily loaded case. *SIAM J. Comput.*, 35(6):1350–1385, 2006.

[3] P. Berenbrink, T. Friedetzky, Z. Hu, and R. Martin. On weighted balls-into-bins games. *Theoret. Comput. Sci.*, 409(3):511–520, 2008.

[4] L. E. Celis, O. Reingold, G. Segev, and U. Wieder. Balls and bins: Smaller hash families and faster evaluation. In R. Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 599–608, 2011.

[5] T. M. Cover and J. A. Thomas. *Elements of information theory (2. ed.).* Wiley, 2006.

[6] A. Czumaj and V. Stemann. Randomized allocation processes. *Random Structures Algorithms*, 18(4):297–331, 2001.

[7] D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[8] R. Dwivedi, O. N. Feldheim, O. Gurel-Gurevich, and A. Ramdas. The power of online thinning in reducing discrepancy. *Probab. Theory Related Fields*, 174(1-2):103–131, 2019.

[9] D. L. Eager, E. D. Lazowska, and J. Zahorjan. Adaptive load sharing in homogeneous distributed systems. *IEEE Transactions on Software Engineering*, SE-12(5):662–675, 1986.

[10] O. N. Feldheim and O. Gurel-Gurevich. The power of thinning in balanced allocation. *Electron. Commun. Probab.*, 26:Paper No. 34, 8, 2021.

[11] O. N. Feldheim, O. Gurel-Gurevich, and J. Li. Long-term balanced allocation via thinning, 2021. arXiv:2110.05009.

[12] O. N. Feldheim and J. Li. Load balancing under d-thinning. *Electron. Commun. Probab.*, 25:Paper No. 1, 13, 2020.

[13] T. Friedrich, M. Gairing, and T. Sauerwald. Quasirandom load balancing. *SIAM J. Comput.*, 41(4):747–771, 2012.

[14] R. Gibbons, F. Kelley, and P. Key. Dynamic alternative routing-modelling and behavior. In *Proceedings of the 12 International Teletraffic Congress, Torino, Italy*. Elsevier, Amsterdam, 1988.

[15] P. B. Godfrey. Balls and bins with structure: balanced allocations on hypergraphs. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 511–517. ACM, New York, 2008.

[16] K. Iwama and A. Kawachi. Approximated two choices in randomized load balancing. In R. Fleischer and G. Trippen, editors, *Algorithms and Computation*, pages 545–557, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[17] R. M. Karp, M. Luby, and F. Meyer auf der Heide. Efficient PRAM simulation on a distributed memory machine. *Algorithmica*, 16(4-5):517–542, 1996.

[18] D. Los and T. Sauerwald. Balanced allocations with incomplete information: The power of two queries, 2021. arXiv:2107.03916.

[19] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: theory of majorization and its applications*. Springer Series in Statistics. Springer, New York, second edition, 2011.

[20] M. Mitzenmacher. On the analysis of randomized load balancing schemes. *Theory Comput. Syst.*, 32(3):361–386, 1999.

[21] M. Mitzenmacher, B. Prabhakar, and D. Shah. Load balancing with memory. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 799–808. IEEE, 2002.

[22] M. Mitzenmacher, A. W. Richa, and R. Sitaraman. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing, Vol. I, II*, volume 9 of *Comb. Optim.*, pages 255–312. Kluwer Acad. Publ., Dordrecht, 2001.

[23] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995.

[24] Y. Peres, K. Talwar, and U. Wieder. Graphical balanced allocations and the $(1+\beta)$-choice process. *Random Structures Algorithms*, 47(4):760–775, 2015.

[25] M. Raab and A. Steger. "Balls into bins"—a simple and tight analysis. In *Proceedings of 2nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM'98)*, volume 1518, pages 159–170. Springer, 1998.

[26] J. Spencer. Balancing games. *J. Combinatorial Theory Ser. B*, 23(1):68–74, 1977.

[27] K. Talwar and U. Wieder. Balanced allocations: the weighted case. In D. S. Johnson and U. Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 256–265. ACM, 2007.

[28] B. Vöcking. How asymmetry helps load balancing. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*, pages 131–141. IEEE Computer Society, 1999.

[29] U. Wieder. Balanced allocations with heterogenous bins. In P. B. Gibbons and C. Scheideler, editors, *SPAA 2007: Proceedings of the 19th Annual ACM Symposium on Parallelism in Algorithms and Architectures, San Diego, California, USA, June 9-11, 2007*, pages 188–193. ACM, 2007.

[30] U. Wieder. Hashing, load balancing and multiple choice. *Found. Trends Theor. Comput. Sci.*, 12(3-4):275–379, 2017.

A Analysis and Concentration Inequalities

Lemma A.1 (Log-Sum Inequality [5, Theorem 2.7.1]). *For any natural number $n \geqslant 1$, let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be nonnegative real numbers and $a := \sum_{i=1}^{n} a_i$ and $b := \sum_{i=1}^{n} b_i$. Then*

$$\sum_{i=1}^{n} a_i \log\left(\frac{a_i}{b_i}\right) \geqslant a \log\left(\frac{a}{b}\right) \quad \text{or equivalently} \quad \sum_{i=1}^{n} a_i \log\left(\frac{b_i}{a_i}\right) \leqslant a \log\left(\frac{b}{a}\right).$$

The following lemma is similar to [13, Lemma A.1.]

Lemma A.2. *Let $(a_k)_{k=1}^n$, $(b_k)_{k=1}^n$ be non-negative and $(c_k)_{k=1}^n$ be non-negative and non-increasing. If $\sum_{k=1}^{i} a_k \leqslant \sum_{k=1}^{i} b_k$ holds for all $1 \leqslant i \leqslant n$ then,*

$$\sum_{k=1}^{n} a_k \cdot c_k \leqslant \sum_{k=1}^{n} b_k \cdot c_k. \tag{A.1}$$

Proof. We shall prove (A.1) holds by induction on $n \geqslant 1$. The base case $n = 1$ follows immediately from the fact that $a_1 \leqslant b_1$ and $c_1 \geqslant 0$. Thus we assume $\sum_{k=1}^{n-1} a_k \cdot c_k \leqslant \sum_{k=1}^{n-1} b_k \cdot c_k$ holds for all sequences $(a_k)_{k=1}^{n-1}$, $(b_k)_{k=1}^{n-1}$ and $(c_k)_{k=1}^{n-1}$ satisfying the conditions of the lemma.

For the inductive step, suppose we are given sequences $(a_k)_{k=1}^n$, $(b_k)_{k=1}^n$ and $(c_k)_{k=1}^n$ satisfying the conditions of the lemma. If $c_2 = 0$ then, since $(c_k)_{k=1}^n$ is non-increasing and non-negative, $c_k = 0$ for all $k \geqslant 2$. Thus as $a_1 \leqslant b_1$ and $c_1 \geqslant 0$ by the precondition of the lemma, we conclude

$$\sum_{k=1}^{n} a_k \cdot c_k = a_1 \cdot c_1 \leqslant b_1 \cdot c_1 = \sum_{k=1}^{n} b_k \cdot c_k.$$

We now treat the case $c_2 > 0$. Define the non-negative sequences $(a'_k)_{k=1}^{n-1}$ and $(b'_k)_{k=1}^{n-1}$ as follows:

- $a'_1 = \frac{c_1}{c_2} \cdot a_1 + a_2$ and $a'_k = a_{k+1}$ for $2 \leqslant k \leqslant n-1$,
- $b'_1 = \frac{c_1}{c_2} \cdot b_1 + b_2$ and $b'_k = b_{k+1}$ for $2 \leqslant k \leqslant n-1$,

Then as the inequalities $c_1 \geqslant c_2$, $a_1 \leqslant b_1$ and $\sum_{i=1}^{n} a_k \leqslant \sum_{i=1}^{n} b_k$ hold by assumption, we have

$$\sum_{k=1}^{n-1} a'_k = \left(\frac{c_1}{c_2} - 1\right) a_1 + \sum_{k=1}^{n} a_k \leqslant \left(\frac{c_1}{c_2} - 1\right) b_1 + \sum_{k=1}^{n} b_k = \sum_{k=1}^{n-1} b'_k.$$

Thus if we also let $(c'_k)_{k=1}^{n-1} = (c_{k+1})_{k=1}^{n-1}$, which is positive and non-increasing, then

$$\sum_{k=1}^{n-1} a'_k \cdot c'_k \leqslant \sum_{k=1}^{n-1} b'_k \cdot c'_k,$$

by the inductive hypothesis. However

$$\sum_{k=1}^{n-1} a'_k \cdot c'_k = \left(\frac{c_1}{c_2} \cdot a_1 + a_2\right) c_2 + \sum_{k=2}^{n-1} a_{k+1} \cdot c_{k+1} = \sum_{k=1}^{n} a_k \cdot c_k,$$

and likewise $\sum_{k=1}^{n-1} b'_k \cdot c'_k = \sum_{k=1}^{n} b_k \cdot c_k$. The result follows. □

For completeness, we define Schur-convexity (see [19]) and state two basic results:

Definition A.3. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is Schur-convex if for any non-decreasing $x, y \in \mathbb{R}^n$, if x majorizes y then $f(x) \geqslant f(y)$. A function f is Schur-concave if $-f$ is Schur-convex.*

Lemma A.4. *Let $g : \mathbb{R} \to \mathbb{R}$ be a convex (resp. concave) function. Then $g(x_1, \ldots, x_n) := \sum_{i=1}^{n} g(x_i)$ is Schur-convex (resp. Schur-concave).*

Lemma A.5. *For any $\alpha > 0$, for any $\beta \in \mathbb{R}$ and any $\Delta \in \mathbb{R}$, consider $f(x_1, x_2, \ldots, x_k) = \sum_{j=1}^{k} \exp(-\alpha x_j)$, where $\sum_{j=1}^{k} x_j \geqslant \Delta$ and $x_j \geqslant \beta$ for all $1 \leqslant j \leqslant k$. Then,*

$$f(x_1, x_2, \ldots, x_k) \leqslant (k-1) \cdot \exp(-\alpha \cdot \beta) + 1 \cdot \exp(-\alpha \cdot (\Delta - (k-1) \cdot \beta)).$$

Proof. Note that by Lemma A.4, it follows that $f(x_1, \ldots, x_n)$ is Schur-concave, as $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} g(x_i)$ and $g(z) = e^{-\alpha z}$ is concave for $\alpha > 0$. As a consequence, the function attains its maximum if the values $(x_1, x_2, \ldots, x_k)$ are as "spread out" as possible, i.e., if any prefix sum of the values ordered non-increasingly is as large as possible. □

Lemma A.6. *Consider a normalized load vector y and vectors y_1, y_2 obtained by adding two balls to y, such that y_2 majorizes y_1. Then, for any $\alpha > 0$,*

$$\sum_{i : y_{1,i} \geqslant 2} e^{\alpha y_{1,i}} \leqslant \sum_{i : y_{2,i} \geqslant 2} e^{\alpha y_{2,i}} + 2e^{3\alpha}.$$

Proof. Let $f(z) = \sum_{i : z_i \geqslant 2} e^{\alpha z_i}$. Note that for any $k \in \mathbb{N}$ and $\mathcal{R} = \{v \in \mathbb{R}^k : v_i \geqslant 2\}$, the function $\widetilde{f} : \mathcal{R} \to \mathbb{R}$ with $\widetilde{f}(z) = \sum_{i=1}^{k} e^{\alpha z_i}$ is Schur-convex, as it is the sum of convex functions $g(u) = e^{\alpha u}$ for $\alpha > 0$ (Lemma A.4).

We consider cases based on the (at most) two positions i_1 and i_2 (with $y_{i_1} \geqslant y_{i_2}$) where $y_{1,i_1} > y_{2,i_1}$ and $y_{1,i_2} > y_{2,i_2}$:

- **Case 1** [$y_{1,i_2} \geqslant 2$ (and $y_{1,i_1} \geqslant 2$)]. In this case, because y_2 majorizes y_1, vectors y_1 and y_2 disagree in positions j with $y_j \geqslant 2$. So by considering $\mathcal{J} := \{j : y_{1,j} \geqslant 2\} = \{j : y_{2,j} \geqslant 2\} =: j_1, \ldots, j_k$, for $k = |\mathcal{J}|$, the function $\widetilde{f} : \mathcal{R} \to \mathbb{R}$ is Schur-convex. Hence,
$$f(y_1) = \widetilde{f}(y_{1,j_1}, \ldots, y_{1,j_k}) \leqslant \widetilde{f}(y_{2,j_1}, \ldots, y_{2,j_k}) = f(y_2).$$

- **Case 2** [$y_{1,i_1} \geqslant 2$ and $y_{1,i_2} < 2$]. The contribution of $y_{1,i_2} < 2$ is at most $e^{3\alpha}$. For y_{1,i_1}, because of y_2 majorizes y_1, the second ball must be placed in j, such that $y_{2,j} \geqslant y_{1,i_1}$. So, $f(y_1) \leqslant f(y_2) + e^{3\alpha}$.

- **Case 3** [$y_{1,i_1} < 2$]. Both can have a maximum contribution of $2e^{3\alpha}$, so $f(y_1) \leqslant f(y_2) + 2e^{3\alpha}$.

Combining the three cases, the claim follows. □

Lemma A.7. *For integers $0 \leqslant r_0 \leqslant r_1$, and real numbers $\varepsilon \in (0, 2/3)$ and $\xi > 0$, we let $f : [r_0, r_1] \cap \mathbb{N} \to [0, 1]$ be a function satisfying*

1. *$f(r_0) \leqslant 1 - \varepsilon$,*
2. *$f(r_1) \geqslant \varepsilon$,*
3. *and for all $t \in [r_0, r_1 - 1]$ we have $f(t+1) \leqslant f(t) + \xi$.*

Then,
$$\left|\left\{ t \in [r_0, r_1] \colon f(t) \in \Big(\frac{\varepsilon}{2}, 1 - \frac{\varepsilon}{2}\Big) \right\}\right| \geqslant \min\left\{ \frac{\varepsilon}{2\xi}, r_1 - r_0 \right\}.$$

Proof. First consider the case where there exists a $t \in [r_0, r_1]$ with $f(t) \geqslant 1-\varepsilon/2$. Further, let t be the first round with that property, hence for any $s \in [r_0, t]$, $f(s) < 1-\varepsilon/2$. Further, thanks to the third property, $f(t-x) \geqslant f(t) - x\xi \geqslant 1-\varepsilon/2 - x\xi$ for any $x \geqslant 0$ (which also implies $r_0 - t \geqslant \varepsilon/(2\xi)$), and thus as long as $0 \leqslant x \leqslant \varepsilon/(2\xi)$,

$$f(t-x) \geqslant f(t) - x\xi \geqslant 1 - \varepsilon/2 - \varepsilon/2 \geqslant \varepsilon/2,$$

since $\varepsilon \leqslant 2/3$. Hence for any $s \in [t - \varepsilon/(2\xi), t]$,

$$f(s) \in [\varepsilon/2, 1-\varepsilon/2].$$

Now consider the case where for all rounds $t \in [r_0, r_1]$ we have $f(t) \leqslant 1-\varepsilon/2$. Since $f(r_1) \geqslant \varepsilon$, we conclude for any $x \leqslant \varepsilon/(2\xi)$,

$$f(r_1 - x) \geqslant f(r_1) - x\xi \geqslant \varepsilon - \varepsilon/2 \geqslant \varepsilon/2.$$

Hence for any $s \in [\max\{r_0, r_1 - \varepsilon/(2\xi)\}, r_1]$ we have $f(s) \in [\varepsilon/2, 1-\varepsilon/2]$. □

Lemma A.8 (Method of Bounded Independent Differences [7, Corollary 5.2]). *Let f be a function of n independent random variables $X_1, \ldots, X_n$, where each X_i takes values in a set A_i. Assume that for each $i \in [n]$ there exists a $c_i \geqslant 0$ such that*

$$\left|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)\right| \leqslant c_i,$$

for any $x_1 \in A_1, \ldots, x_{i-1} \in A_{i-1}, x_i, x_i' \in A_i, x_{i+1} \in A_{i+1}, \ldots, x_n \in A_n$. Then for any real $t > 0$

$$\mathbf{Pr}\left[\, f < \mathbf{E}\left[\, f \,\right] - t \,\right], \mathbf{Pr}\left[\, f > \mathbf{E}\left[\, f \,\right] + t \,\right] \leqslant \exp\left(\frac{-t^2}{2\sum_{i=1}^n c_i^2}\right).$$

Lemma A.9 (Azuma's Inequality for Super-Martingales [7, Problem 6.5]). *Let $X_0, \ldots, X_n$ be a super-martingale satisfying $|X_i - X_{i-1}| \leqslant c_i$ for any $i \in [n]$, then for any $\lambda > 0$,*

$$\mathbf{Pr}\left[\, X_n \geqslant X_0 + \lambda \,\right] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

Claim A.10. *For any $\varepsilon \in (0,1)$, we have $(1-\frac{\varepsilon}{2})/(1-\varepsilon) \geqslant 1 + \frac{\varepsilon}{2}$.*

Proof. For $\varepsilon \in (0,1)$ the following chain of double implications holds

$$\frac{1-\frac{\varepsilon}{2}}{1-\varepsilon} \geqslant 1 + \frac{\varepsilon}{2} \Leftrightarrow 1 - \frac{\varepsilon}{2} \geqslant (1-\varepsilon)\cdot\left(1+\frac{\varepsilon}{2}\right) = 1 - \frac{\varepsilon}{2} + \frac{\varepsilon^2}{2} \Leftrightarrow \frac{\varepsilon^2}{2} \geqslant 0.$$ □

Lemma A.11 ([23, Proposition B.3]). *For any integer $n \geqslant 1$ and real $|x| \leqslant n$ we have $(1+x/n)^n \geqslant e^x(1-x^2/n)$.*

Lemma A.12. *Consider any sequence $(z_i)_{i\in\mathbb{N}}$ such that, for some $a > 0$ and $b > 0$, for every $i \geqslant 1$,*

$$z_i \leqslant z_{i-1} \cdot a + b.$$

Then for every $i \in \mathbb{N}$,

$$z_i \leqslant z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j.$$

Further, if $a < 1$,

$$z_i \leqslant z_0 \cdot a^i + \frac{b}{1-a}.$$

Proof. We will prove the first claim by induction. For $i = 0$, $z_0 \leqslant z_0$. Assume the induction hypothesis holds for some $i \geqslant 0$, then since $a > 0$,

$$z_{i+1} \leqslant z_i \cdot a + b \leqslant \left(z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j\right) \cdot a + b = z_0 \cdot a^{i+1} + b \cdot \sum_{j=0}^{i} a^j.$$

Hence, the first claim follows. The second part of the claim is immediate, since for $a \in (0,1)$, $\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$. □

B Counterexamples for Exponential Potential Functions

In this section, we present two configurations for which the potential functions Φ of Section 5 and Λ of Section 8.2 may increase in expectation over one round, if we don't condition on "good events".

Claim B.1. *For any constant $\alpha > 0$ and for sufficiently large n, consider the (normalized) load configuration,*

$$y^t = (\sqrt{n}, \underbrace{0, \ldots, 0}_{n-\sqrt{n}-1 \text{ bins}}, \underbrace{-1, \ldots, -1}_{\sqrt{n} \text{ bins}}).$$

Then, for the Packing *process, the potential function $\Phi^t := \sum_{i:y_i^t \geqslant 0} e^{\alpha y_i^t}$ will increase in expectation, i.e.,*

$$\mathbf{E}[\, \Phi^{t+1} \mid \mathfrak{F}^t \,] \geqslant \Phi^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big).$$

Proof. Consider the contribution of bin $i = 1$, with $y_1^t = \sqrt{n}$.

$$\mathbf{E}[\, \Phi_1^{t+1} \mid \mathfrak{F}^t \,] = e^{\alpha\sqrt{n}} \cdot \Big(1 + \frac{1}{n} \cdot (e^{\alpha - \alpha/n} - 1) + \frac{n - \sqrt{n} - 1}{n} \cdot (e^{-\alpha/n} - 1) + \frac{\sqrt{n}}{n} \cdot (e^{-2\alpha/n} - 1)\Big).$$

Now using a Taylor estimate $e^z \geqslant 1 + z + 0.3z^2$ for $z \geqslant -1.5$,

$$\begin{aligned}
\mathbf{E}[\, \Phi_1^{t+1} \mid \mathfrak{F}^t \,] &\geqslant e^{\alpha\sqrt{n}} \cdot \Big(1 + \frac{1}{n} \cdot \Big(\alpha - \frac{\alpha}{n} + 0.3 \cdot \Big(\alpha - \frac{\alpha}{n}\Big)^2\Big) + \frac{n - \sqrt{n} - 1}{n} \cdot \Big(-\frac{\alpha}{n} + 0.3 \cdot \frac{\alpha^2}{n^2}\Big) \\
&\qquad + \frac{\sqrt{n}}{n} \cdot \Big(-\frac{2\alpha}{n} + 1.2 \cdot \frac{\alpha^2}{n^2}\Big)\Big) \\
&= e^{\alpha\sqrt{n}} \cdot \Big(1 + \frac{\alpha + 0.3 \cdot \alpha^2}{n} - \frac{\alpha}{n} + o(n^{-1})\Big) \\
&= e^{\alpha\sqrt{n}} \cdot \Big(1 + 0.3 \cdot \frac{\alpha^2}{n} + o(n^{-1})\Big) \\
&\geqslant e^{\alpha\sqrt{n}} \cdot \Big(1 + 0.2 \cdot \frac{\alpha^2}{n}\Big).
\end{aligned}$$

At round t, the contribution of the rest of the bins is at most n. Note that for sufficiently large n, we have $n \cdot (1 + 0.1 \cdot \frac{\alpha^2}{n}) < 0.1 \cdot \alpha^2 \cdot e^{\alpha\sqrt{n}}$. Hence,

$$\begin{aligned}
\mathbf{E}[\, \Phi^{t+1} \mid \mathfrak{F}^t \,] &\geqslant e^{\alpha\sqrt{n}} \cdot \Big(1 + 0.2 \cdot \frac{\alpha^2}{n}\Big) \geqslant e^{\alpha\sqrt{n}} \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) + 0.1 \cdot \frac{\alpha^2}{n} \cdot e^{\alpha\sqrt{n}} \\
&\geqslant \Phi_1^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) + n \cdot (1 + 0.1 \cdot \frac{\alpha^2}{n}) \\
&\geqslant \Phi_1^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) + \Bigg(\sum_{i>1, y_i^t \geqslant 0} \Phi^t\Bigg) \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) \\
&= \Phi^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big).
\end{aligned}$$

□

In contrast to processes with constant probability bias in every round, such as those studied in [24], for the Mean-Thinning process, there exists configurations where the potential Λ for constant α will increase in expectation over a single round, even when it is $\omega(n)$. This is because in the worst-case, the Mean-Thinning process may have a $\Theta(1/n)$ probability bias, corresponding to Towards-Min process in [24].

Claim B.2. *For any constant $\alpha > 0$ and for sufficiently large n consider the (normalized) load configuration,*

$$y^t = \Big(n^2, n, n, \ldots, n, -\frac{n \cdot (2n-3)}{2}, -\frac{n \cdot (2n-3)}{2}\Big).$$

Then for the Mean-Thinning *process, the exponential potential Λ will increase in expectation over one step, i.e.,*

$$\mathbf{E}[\, \Lambda^{t+1} \mid \mathfrak{F}^t \,] \geqslant \Lambda^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big),$$

and $\delta^s \geqslant 1 - 2/n$ for $s \in [t, t+n^2)$.

Proof. Consider the contribution of bin $i = 1$, with $y_i^t = n^2$. The probability of allocating to that bin is $\frac{1-\frac{2}{n}}{n}$, so using the Taylor estimate $e^z \geqslant 1 + z + 0.3 \cdot z^2$ for $z > 1.5$,

$$\begin{aligned}
\mathbf{E}[\, \Lambda_1^{t+1} \mid \mathfrak{F}^t \,] &= e^{\alpha n^2} \cdot e^{-\alpha/n} \cdot (1 + (e^\alpha - 1) \cdot \frac{1 - \frac{2}{n}}{n}) \\
&\geqslant e^{\alpha n^2} \cdot (1 - \frac{\alpha}{n} + 0.3 \cdot \frac{\alpha^2}{n^2}) \cdot \Big(1 + (\alpha + 0.3 \cdot \alpha^2) \cdot \frac{1 - \frac{2}{n}}{n}\Big) \\
&= e^{\alpha n^2} \cdot \Big(1 - \frac{\alpha}{n} + (\alpha + 0.3\alpha^2) \cdot \frac{1}{n} + o(n^{-1})\Big) \\
&= e^{\alpha n^2} \cdot \Big(1 + 0.2 \cdot \frac{\alpha^2}{n}\Big).
\end{aligned}$$

Now note that for sufficiently large n,

$$0.1 \cdot \frac{\alpha^2}{n} \cdot e^{\alpha n^2} \geqslant \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) \cdot n \cdot e^{\alpha n^2} \cdot e^{-\alpha \cdot 3n/2} = \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) \cdot n \cdot e^{\alpha n(2n-3)/2} \geqslant \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) \cdot \Big(\sum_{i>1} \Lambda_i^t\Big),$$

since $\Lambda_i^t \leqslant e^{\alpha n(2n-3)/2}$ for $i > 1$. Hence,

$$\begin{aligned}
\mathbf{E}[\, \Lambda^{t+1} \mid \mathfrak{F}^t \,] \geqslant \mathbf{E}[\, \Lambda_1^{t+1} \mid \mathfrak{F}^t \,] &\geqslant e^{\alpha n^2} \cdot \Big(1 + 0.2 \cdot \frac{\alpha^2}{n}\Big) \\
&= e^{\alpha n^2} \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) + 0.1 \cdot \frac{\alpha^2}{n} \cdot e^{\alpha n^2} \\
&\geqslant \Lambda_1^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) + \Big(\sum_{i>1} \Lambda_i^t\Big) \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big) \\
&= \Lambda^t \cdot \Big(1 + 0.1 \cdot \frac{\alpha^2}{n}\Big).
\end{aligned}$$

Finally, since there are $n - 2$ overloaded bins with overload at least n and these can decrease by at most $1/n$ in each round, we have that $\delta^s \geqslant 1 - 2/n$ for any $s \in [t, t+n^2)$. $\square$

C Experimental Results

In this section, we present some empirical results for four processes: Mean-Thinning, Twinning, Packing and Caching (Table 2 and Fig. 7) and compare their load with that of a $(1+\beta)$ process with $\beta = 0.5$ and with that of Two-Choice process.

n	Mean-Thinning	Twinning	Packing	Caching
10^3	**4** : 2% **5** : 38% **6** : 35% **7** : 15% **8** : 8% **9** : 1% **12** : 1%	**8** : 3% **9** : 21% **10** : 25% **11** : 18% **12** : 13% **13** : 8% **14** : 9% **17** : 3%	**6** : 3% **7** : 14% **8** : 30% **9** : 23% **10** : 15% **11** : 8% **12** : 4% **13** : 1% **14** : 1% **15** : 1%	**2** : 67% **3** : 33%
10^4	**6** : 2% **7** : 30% **8** : 38% **9** : 19% **10** : 9% **11** : 1% **14** : 1%	**11** : 1% **12** : 9% **13** : 24% **14** : 22% **15** : 13% **16** : 9% **17** : 8% **18** : 5% **19** : 6% **20** : 1% **21** : 1% **26** : 1%	**9** : 2% **10** : 17% **11** : 28% **12** : 14% **13** : 22% **14** : 11% **15** : 3% **16** : 2% **17** : 1%	**2** : 5% **3** : 95%
10^5	**8** : 3% **9** : 32% **10** : 38% **11** : 15% **12** : 6% **13** : 3% **14** : 3%	**14** : 2% **15** : 5% **16** : 25% **17** : 28% **18** : 17% **19** : 10% **20** : 8% **21** : 1% **22** : 1% **23** : 3%	**12** : 2% **13** : 16% **14** : 20% **15** : 28% **16** : 23% **17** : 5% **18** : 3% **19** : 1% **20** : 2%	**3** : 100%

Table 2: Summary of observed gaps for $n \in \{10^3, 10^4, 10^5\}$ bins and $m = 1000 \cdot n$ number of balls, for 100 repetitions. The observed gaps are in bold and next to that is the % of runs where this was observed.

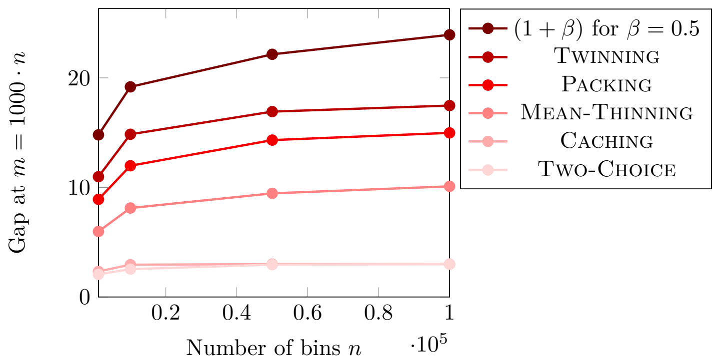

Figure 7: Average Gap vs. $n \in \{10^3, 10^4, 5 \cdot 10^4, 10^5\}$ for the experimental setup of Table 2.